DOCUMENT RESUME

ED 385 572 TM 024 010

AUTHOR McPeek, W. Miles; Wild, Cheryl L.
TITLE Identifying Differentially Functioning Items in the NTE Core Battery.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-92-62
PUB DATE Oct 92
NOTE 116p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Beginning Teachers; Blacks; Elementary Secondary Education; Evaluation Methods; Females; Hispanic Americans; *Identification; *Item Bias; Males; Racial Differences; Sex Differences; *Teacher Evaluation; *Test Items; Whites
IDENTIFIERS *Mantel Haenszel Procedure; *NTE Core Battery; Research Replication

ABSTRACT

The use of the Mantel-Haenszel statistic was investigated as a methodology for identifying differentially functioning items on the NTE Programs Core Battery. Retrospective analyses of the data collected over a 3-year period are reported for Black/White, Hispanic/White, and female/male comparisons in 50 samples ranging from 88 to 23,773 teacher candidates. The samples cover the three tests that comprise the Core Battery: Communication Skills, General Knowledge, and Professional Knowledge. Replicability of differential item functioning values from two administrations based on correlational data was modest to high, with the greatest replicability found in the largest samples. Results were less consistent when only the items identified as being differentially difficult were considered. Generally, 6 to 9% of the items were identified as being differentially difficult per comparison, with a greater number of items flagged in female/male analyses than in Black/White or Hispanic/White analyses. There was little overlap in the items flagged among the three different analyses. Due to the small number of items studied and the post hoc nature of the sample mean the findings should be considered exploratory. Appendix A contains item classifications, and Appendix B contains eight supplemental tables. (Contains 18 tables and 22 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-92-62

ED 385 572

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

IDENTIFYING DIFFERENTIALLY FUNCTIONING ITEMS IN THE NTE CORE BATTERY

W. Miles McPeek
Cheryl L. Wild

Educational Testing Service
Princeton, New Jersey
October 1992

TM024010

BEST COPY AVAILABLE

Identifying Differentially Functioning Items in the NTE Core Battery

Prepared by

W. Miles McPeek

and

Cheryl L. Wild

Educational Testing Services

July 1992

Copyright © 1992. Educational Testing Service. All rights reserved.

Table of Contents

	Page
List of Tables	3
List of Contents Appendix A	5
List of Tables Appendix B	6
Acknowledgments	7
Abstract	8
Introduction	9
Procedures	10
Results	16
Conclusions	39
References	43
Appendix A	A1
Appendix B	B1

LIST OF TABLES

		Page
Table 1	Summary of Sample Sizes by Test Form and Administration Date	11
Table 2	Summary of Mean Subscore Differences between Black and White Groups in Standard Deviation Units	12
Table 3	Summary of Mean Subscore Differences between Hispanic and White Groups in Standard Deviation Units	13
Table 4	Summary of Mean Subscore Differences between Female and Male Groups in Standard Deviation Units	14
Table 5	Average DIF Values for Black and White Groups	18
Table 6	Average DIF values for Hispanic and White Groups	19
Table 7	Average DIF values for Female and Male Groups	20
Table 8	Correlations of DIF Index Values on the Same Form at Two Administrations of the NTE Core Battery	21
Table 9	Agreement of Flagged NTE Questions on the Same Form at Two Administrations: Black/White Comparison	22
Table 10	Agreement of Flagged NTE Questions on the Same Form at Two Administrations: Hispanic/White Comparison	23
Table 11	Agreement of Flagged NTE Questions on the Same Form at Two Administrations: Female/Male Comparison	24
Table 12	Number and Percentage of Items Flagged in the Black/White Analyses	26
Table 13	Number and Percentage of Items Flagged in the Hispanic/White Analyses	27
Table 14	Number and Percentage of Items Flagged in the Female/Male Analyses	28

LIST OF TABLES (continued)

Page

Table 15 Summary of NTE Variables Studied 31

Table 16 Summary of Communication Skills Variables Significant at the .05 Level and Beyond 32

Table 17 Summary of General Knowledge Variables Significant at the .05 Level and Beyond 33

Table 18 Summary of Professional Knowledge Variables Significant at the .05 Level and Beyond 34

List of Contents APPENDIX A - Item Classifications

Page

Introduction A1

Summary of Variables on Which Items Were Classified A2

Variables Common to All Tests A4

Variables Studied in Communication Skills: Listening A5

Variables Studied in Communication Skills: Reading A8

Variables Studied in Communication Skills: Writing A12

Variables Studied in General Knowledge: Literature and Fine Arts . A15

Variables Studied in General Knowledge: Mathematics A18

Variables Studied in General Knowledge: Science A24

Variables Studied in General Knowledge: Social Studies A30

Variables Studied in Professional Knowledge A32

LIST OF TABLES APPENDIX B

		Page
Table B1	Means and Standard Deviations of Categories for Variables on NTE Communication Skills: Reading	B1
Table B2	Means and Standard Deviations of Categories for Variables on NTE Communication Skills: Listening . . .	B6
Table B3	Means and Standard Deviations of Categories for Variables on NTE Communication Skills: Writing	B10
Table B4	Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Social Studies . .	B14
Table B5	Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Mathematics . . .	B18
Table B6	Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Science	B21
Table B7	Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Literature and Fine Arts	B24
Table B8	Means and Standard Deviations of Categories for Variables on NTE Professional Knowledge	B28

Acknowledgments

The authors would like to thank the numerous individuals who contributed to the design and implementation of this study. Special thanks should go to Ruth Austria, without whose help we could not have completed this project. Ruth coordinated the coding of the item classifications, keying of the data, organization of the output, preparation of the tables, proofing of the report, and numerous other administrative activities.

Paul Holland also made significant contributions to this report by suggesting the data analyses, doing the basic theoretical work to allow us to use the Mantel-Haenszel statistic, reviewing the data analyses and reviewing the report. Minhwei Wang and Dorothy Thayer conducted the data analyses. Aileen Cramer, Frank Fornoff, Miriam Levin, Lane MacNeish, William Eric Perkins, Leslie Pierce, Alexandra Radbil, Paul Ramsey, Edith Robinson, and Charlotte Solomon helped in developing the hypotheses and coding the data. Nancy Parr made numerous, significant editorial contributions. Numerous other individuals consulted with us at various steps of this project and, although unlisted here, also deserve our thanks.

Abstract

The purpose of this study is to evaluate the use of the Mantel-Haenszel statistic as a methodology for identifying differentially functioning items on the NTE Programs Core Battery tests. Retrospective analyses of data collected over a three-year period are reported for Black/White, Hispanic/White, and female/male comparisons. The three tests that comprise the Core Battery are included in the analyses--Communication Skills, General Knowledge, and Professional Knowledge. Data concerning the replicability of the results, the number of items identified as differentially difficult, and the characteristics of the test questions associated with differential difficulty are discussed. Replicability of DIF values from two administrations based on correlational data was modest to high, with greatest replicability found in the largest samples. Although the replicability is modest to high when considering all the items, the results are less consistent when only the items identified as showing differential difficulty are considered. Generally, 6 to 9 percent of the items were identified as being differentially difficult per comparison, with a greater number of items flagged in the female/male analyses than in the Black/White or Hispanic/White analyses. There was little overlap among the items flagged in the three different analyses. Although analyses suggested some content characteristics that may be related to differential item functioning, the findings should be considered exploratory due to the small number of items studied and the post hoc nature of the sample. Also, the findings should not be assumed to apply to populations other than NTE test takers.

Identifying Differentially Functioning Items in the NTE Core Battery Tests

Differential item functioning (DIF) is said to exist when two groups of people, matched in terms of their relevant knowledge and skill, perform differently on an item. In the summer of 1986, ETS began to introduce a statistical measure of DIF as an additional routine tool in the test development process. The introduction of this tool was the culmination of a series of studies to identify a statistical methodology that could be used in a variety of testing programs to detect test questions exhibiting differential item functioning among various subgroups. This paper presents results of one of the studies--the study investigating the use of the Mantel-Haenszel statistic on retrospective data from NTE Programs.

A coordinated research program, begun in 1983, evaluated the Mantel-Haenszel statistic as a method to investigate DIF. Papers by Holland (1985), Holland and Thayer (1986), and Phillips and Holland (1986) present the statistical evaluations and development of the procedures. The procedures were then used in retrospective analyses of the NTE Core Battery tests, the Graduate Record Examinations General Test, the Graduate Management Admission Test, the Multistate Insurance Licensing Test, and the Scholastic Aptitude Test. Parts of these retrospective studies have been reported in several places (Carlton & Harris, 1989; McPeek & Wild, 1986, 1987; O'Neill, Wild, & McPeek, 1989; Wild & McPeek, 1986).

Numerous research projects at ETS have investigated differential item functioning over the years. These include applications of delta plot, analyses of variance, standardization, and item response theory techniques (e.g., Angoff & Ford, 1973; Cardall & Coffman, 1964; Dorans & Kulick, 1986; Lord, 1980). The earlier studies differ from the effort reported here mainly in the statistical procedure used (the Mantel-Haenszel) and the focus on developing procedures that can be used operationally. Operational procedures that are currently in use at ETS are described by Zieky (1988).

Although originally used by biostatisticians (Mantel & Haenszel, 1959) the Mantel-Haenszel procedure was recently proposed by Holland as a statistically powerful and efficient index for identifying differentially functioning test items in operational test forms (Holland, 1985). The Mantel-Haenszel procedure has been compared to other procedures intended to identify differential item performance (Holland & Thayer, 1986; Lord, 1980; Marascuilo & Slaughter, 1981; Mellenberg, 1982; Scheuneman, 1979) and has been recommended for use instead of the item response theory (IRT) methods because it is independent of the assumptions underlying IRT models, is applicable to small samples, and is inexpensive.

The purpose of this study is to evaluate use of the Mantel-Haenszel statistic as a methodology for identifying differentially functioning items on the NTE Core Battery. The NTE Core Battery tests

are designed to measure the academic achievement and basic proficiency of students completing teacher education programs. The battery consists of three separate two-hour tests--the Test of Communication Skills, the Test of General Knowledge, and the Test of Professional Knowledge--which assess knowledge and skills deemed necessary or highly desirable for beginning teachers. The test content specifications are designed to ensure that test takers can demonstrate their basic skills in the areas of listening, reading, and writing in the Communications Skills test, and their knowledge and understanding of the interrelationships of literature and fine arts, mathematics, science, and social studies in the Test of General Knowledge. The Test of Professional Knowledge assesses test takers' understanding of the knowledge and skills that a beginning teacher uses in decision making, with emphasis on the context and process of teaching. The detailed content specifications for these tests are set by committees of experienced teachers and teacher educators. Job analyses and content validity studies relate test content directly to teacher training programs and to job-related skills and knowledge. This study is a retrospective look at data that were collected over a three-year period, 1983-1985, at regular NTE test administrations. Three basic questions will be discussed based on these data:

(1) Are the same questions identified as showing differential item functioning in analyses of test takers from two different administrations of the same test?

(2) What percentages of questions are identified as differentially difficult?

(3) What are some of the characteristics of the test questions that are associated with differential difficulty?

Procedures

Two forms each of the NTE Communication Skills test, General Knowledge test, and Professional Knowledge test were analyzed for this study. Sample sizes and administration dates are summarized in Table 1. Data presented in the response to Question 1 (replicability of results) are based on the October 1983 and March 1985 administrations of Communication Skills Form 1, General Knowledge Form 1, and Professional Knowledge Form 2 and on the October 1983 and March 1984 administrations of Professional Knowledge Form 1. Data presented in the response to Questions 2 and 3 are based on both forms of each test but only on the later administration dates (March 1984 and 1985), which contain larger samples of examinees. Tables 2, 3, and 4 present the mean subscore differences between Black and White, Hispanic and White, and male and female test takers in standard deviation units for each of the different parts of the Communication Skills test and the General Knowledge test, and for the total score of the Professional Knowledge test. The subparts of the Communication Skills test are listening, reading, and writing; the subparts of the General Knowledge Test are social studies, literature and fine arts, mathematics, and science.

Table 1

Summary of Sample Sizes by
Test Form and Administration Date

	Form	Admin. Date	N Male Group	N Female Group	N White Group	N Black Group	N Hispanic Group
Communication Skills	1	10/83	1,625	6,604	6,759	894	190
	2	3/84	3,054	12,029	12,032	1,811	437
	1	3/85	5,398	20,278	21,244	2,174	591
General Knowledge	1	10/83	1,862	7,864	7,975	999	284
	2	3/84	3,250	13,526	13,534	1,884	564
	1	3/85	5,514	23,100	23,773	2,308	839
Professional Knowledge	1	10/83	840	3,462	3,663	440	88
	2	10/83	870	3,418	3,642	439	89
	1	3/84	3,302	12,681	13,044	2,032	415
	2	3/85	5,204	19,125	20,124	2,352	544

Table 2

Summary of Mean Subscore Differences between
Black and White Groups in Standard Deviation Units*

	Maximum Number of Items**	Subscore Differences by Administration and Form		
		10/83	3/84	3/85
Communication Skills		(Form 1)	(Form 2)	(Form 1)
Listening	40	1.30	1.26	1.41
Reading	30	1.30	1.26	1.37
Writing	45	1.11	1.18	1.13
General Knowledge				
Social Studies	30	1.13	1.20	1.09
Mathematics	25	1.30	1.29	1.32
Literature & Fine Arts	35	1.28	1.22	1.26
Science	30	1.34	1.29	1.35
Professional Knowledge				
		(Form 1)	(Form 1)	(Form 2)
	105	1.42	1.42	1.41
		(Form 2)		
	105	1.37		

*Calculated by the formula

$$\frac{\text{(Mean Score White Examinees - Mean Score Black Examinees)}}{\text{Total Group Standard Deviation}}$$

**Occasionally an item is dropped from scoring.

Table 3

Summary of Mean Subscore Differences between
Hispanic and White Groups in Standard Deviation Units*

	Maximum Number of Items**	Subscore Differences by Administration and Form		
		10/83	3/84	3/85
Communication Skills		(Form 1)	(Form 2)	(Form 1)
Listening	40	.82	.75	.91
Reading	30	.66	.56	.80
Writing	45	.67	.65	.82
General Knowledge				
Social Studies	30	.38	.52	.59
Mathematics	25	.52	.65	.81
Literature & Fine Arts	35	.55	.53	.71
Science	30	.45	.55	.70
Professional Knowledge				
		(Form 1)	(Form 1)	(Form 2)
	105	.69	.59	.75
		(Form 2)		
	105	.60		

*Calculated by the formula

$$\frac{\text{(Mean Score White Examinees - Mean Score Hispanic Examinees)}}{\text{Total Group Standard Deviation}}$$

**Occasionally an item is dropped from scoring.

Table 4

Summary of Mean Subscore Differences between
Female and Male Groups in Standard Deviation Units*

	Maximum Number of Items**	Subscore Differences by Administration and Form		
		10/83	3/84	3/85
Communication Skills		(Form 1)	(Form 2)	(Form 1)
Listening	40	-.19	-.11	-.17
Reading	30	-.10	-.09	-.05
Writing	45	-.23	-.27	-.20
General Knowledge				
Social Studies	30	.41	.39	.48
Mathematics	25	.37	.31	.38
Literature & Fine Arts	35	-.09	-.08	-.03
Science	30	.32	.28	.36
Professional Knowledge				
		(Form 1)	(Form 1)	(Form 2)
	105	-.34	-.27	-.16
		(Form 2)		
	105	-.16		

*Calculated by the formula

$$\frac{\text{(Mean Score Male Examinees - Mean Score Female Examinees)}}{\text{Total Group Standard Deviation}}$$

**Occasionally an item is dropped from scoring.

The Mantel-Haenszel odds ratio was calculated using as a matching criterion all of the available items testing the same skill or knowledge. In the Communication Skills test, for example, the Mantel-Haenszel odds ratio for listening items was calculated using all listening items (including the studied item) as the criterion. The matching criterion for the reading items was the score on all reading items; the matching criterion for the writing items was the score on all writing items. Similarly, the four subparts of the General Knowledge test were used as matching criteria. The total score of the Professional Knowledge test was the matching criterion for the items on that test. The total test scores on Communication Skills and General Knowledge were considered as the matching criteria and rejected due to the multidimensional nature of these tests. In order to identify differential item functioning using the Mantel-Haenszel procedure, examinees must be matched on the relevant skill and knowledge tested by the test question. To accomplish this, examinees at each score level are analyzed separately and the results summarized across score levels (Holland & Thayer, 1986). Relatively unidimensional scores provide better matching of the two groups and are therefore preferable.

The odds ratio was then transformed to a scale more familiar to ETS test developers--the difference between the item difficulties of the two matched groups in delta units (deltas have a mean of 13 and a standard deviation of 4, with more difficult items corresponding to larger numbers). We have called this index of differential item difficulty "DIF" in the tables and text. In operational test development practice, a difference in equated deltas greater than one would be very unusual from repeated administrations of the same question. In this study, an absolute value of DIF equal to or greater than one is used as the flagging criterion. (This corresponds to a difference between groups in percent correct of approximately .10 for questions with percent correct values between .25 and .75.) Flagged questions are separated in the summary tables into those questions on which reference group test takers (White or male) performed better than focal group test takers (Black, Hispanic, or female) with comparable test scores, and those questions for which the reverse held. Summaries are presented by skill or area of knowledge to facilitate interpretation.

The replicability of the results (Question 1) is evaluated in two ways. First, the correlations between the DIF values obtained for two administrations are computed; second, the extent to which the same items are consistently classified as having absolute values of DIF equal to or greater than one in each of the two administrations is determined. The correlational analyses provide estimates of the extent to which the orderings of the items by DIF values were similar on data from different administrations. The consistency analyses of flagged items look only at a subset of the items and thus may provide a somewhat different picture. Correlation and consistency indices are calculated separately for Black/White, Hispanic/White, and female/male comparisons.

To study Question 3, we hypothesized variables that might result in differential performance by group. The items were classified according to each variable, and F-tests were obtained to determine if there were differences in mean DIF values for the categories within each variable. For example, "minority stimulus" was one of the hypothesized variables. Items were categorized as follows:

(1) Stimulus material refers to Black people.
(2) Stimulus material refers to Hispanic people.
(3) Stimulus material refers to other minorities.
(4) Stimulus material refers to people of no specified ethnic origin.
(5) Stimulus material does not refer to people.

A listing of all the hypothesized variables and the coding procedures is presented in Appendix A. Because items were classified on about 40 variables, many within-variable categories were small or empty, and there was great overlap among some variables. In this paper we discuss variables that have significant differences at the .01 level and have at least five items in the relevant categories.

Results

This section provides a brief overview of the data and then presents the results organized under the three basic questions of the study.

As Table 1 shows, the Black/White and female/male analyses are based on relatively large sample sizes, whereas in the Hispanic/White comparisons, six of the nine analyses are based on fewer than 500 Hispanic test takers, and two of the analyses have fewer than 100 Hispanic test takers. As will be discussed below, the replicability of DIF values based on analyses of sample sizes with fewer than 500 test takers appears to require additional caution in the interpretation of the results. Since the analysis of these data, formulas for calculating the standard error have been developed (Phillips & Holland, 1986). Tables 2, 3, and 4 present the mean subscore differences in standard-deviation units for the groups being studied. The mean subscore differences are largest for the Black/White comparisons, with White test takers having higher mean subscores than Black test takers. The Hispanic/White standardized mean differences are somewhat smaller than the Black/White differences, with the White test takers again having the higher mean subscores. The female/male standardized mean differences are closer to zero than are those for the Black/White and Hispanic/White comparisons. On some tests women have higher scores, while on other tests men have higher scores.

Tables 5, 6, and 7 present the average DIF values (i.e., mean differences between the reference and focal groups when individuals are matched on items measuring the same skill or subject area) for the separate skills and topics. As expected, all of these mean values are close to zero and differ slightly from zero only due to small effects associated with the nonlinearity of the Mantel-Haenszel weighing procedure for combining data across score levels.

Question 1: Are the same questions identified as showing differential item functioning in analyses of test takers from two different administrations of the same test?

The NTE Communication Skills test Form 1, General Knowledge test Form 1, and Professional Knowledge test Form 2 were given in October 1983 and March 1985. The Professional Knowledge test Form 1 was given in October 1983 and March 1984. Subgroup sample sizes for each administration are given in Table 1. Correlations between the DIF index values for pairs of administrations are given in Table 8, and the percent agreement of flagged items are presented in Tables 9 through 11. These statistics illustrate the extent to which the same items are identified using the DIF index separately for two administrations.

The correlations of the DIF index values (see Table 8) for the female/male samples range from a low of .89 for the reading section to a high of .97 for the science section. In the smaller Black/White samples, the correlations range from a low of .71 in the reading section to a high of .93 in the social studies and literature and fine arts sections. In the Hispanic/White samples, the correlations range from a low of .37 in the reading section to a high of .87 in the social studies section. (In all three analyses, the reading section has a relatively small standard deviation of DIF values that may account for the lower correlations.)

One would expect that the highest correlations would be found in the analyses of the largest sample sizes, and that the correlations would decrease as the sample sizes decrease. The data confirm this expectation. The highest correlations are found, as expected, in the female/male analyses, where the smaller subgroup sample sizes on which DIF values are calculated range from 840 to 5,514. The lowest correlations are found, as expected, in the Hispanic/White analyses, where the smaller subgroup sample sizes on which DIF values are calculated range from 88 to 839.

The correlational analyses give a somewhat more positive picture of the replicability of results than do the analyses of the consistency of the identification of items with absolute values of DIF equal to or greater than one. In the test development process, we want to focus our attention on those items with moderate to large absolute values of DIF. The correlational analyses look at all items, rather than the subset in which we are most interested. The consistency analyses, reported next, focus on the items in which we have the greatest interest.

Table 5

Average DIF* Values for Black and White Groups

		DIF	
		3/84	3/85
Communication Skills			
Listening (40 items)	Mean	.01	-.01
	SD	.64	.51
Reading (30 items)	Mean	-.03	-.01
	SD	.53	.40
Writing (45 items)	Mean	.00	.00**
	SD	.63	.58**
TOTAL	Mean	.00	-.01
	SD	.61	.51
General Knowledge			
Social Studies (30 items)	Mean	-.06**	-.05
	SD	.90**	.83
Mathematics (25 items)	Mean	.03	.03
	SD	.57	.47
Literature & Fine Arts (35 items)	Mean	-.01	-.02**
	SD	.60	.83**
Science (30 items)	Mean	-.01	-.02
	SD	.57	.58
TOTAL	Mean	-.01	-.02
	SD	.49	.50
Professional Knowledge			
(105 items)	Mean	-.02	-.01
	SD	.49	.50

* Negative DIF values indicate that White examinees perform better than Black examinees with comparable test scores; positive DIF values indicate the reverse.

** One item not scored.

Table 6

Average DIF* Values for Hispanic and White Groups

		DIF	
		3/84	3/85
Communication Skills			
Listening (40 items)	Mean	-.02	-.06
	SD	.52	.50
Reading (30 items)	Mean	-.01	-.02
	SD	.36	.52
Writing (45 items)	Mean	-.01	.01**
	SD	.48	.52**
TOTAL	Mean	-.01	-.02
	SD	.46	.51
General Knowledge			
Social Studies (30 items)	Mean	-.01**	-.03
	SD	.65**	.64
Mathematics (25 items)	Mean	-.01	-.01
	SD	.43	.54
Literature & Fine Arts (35 items)	Mean	-.04	-.00**
	SD	.54	.36**
Science (30 items)	Mean	-.03	-.01
	SD	.54	.48
TOTAL	Mean	-.01	-.01
	SD	.50	.50
Professional Knowledge			
(105 items)	Mean	-.00	-.01
	SD	.50	.61

* Negative DIF values indicate that White examinees perform better than Hispanic examinees with comparable test scores; positive DIF values indicate the reverse.

** One item not scored.

Table 7

Average DIF* Values for Female and Male Groups

		DIF	
		3/84	3/85
Communication Skills			
Listening (40 items)	Mean	.05	-.03
	SD	.60	.68
Reading (30 items)	Mean	.03	.03
	SD	.33	.42
Writing (45 items)	Mean	.01	.02**
	SD	.48	.52**
TOTAL	Mean	.03	.00
	SD	.50	.56
General Knowledge			
Social Studies (30 items)	Mean	-.00**	.02
	SD	.62**	.58
Mathematics (25 items)	Mean	.01	-.00
	SD	.63	.66
Literature & Fine Arts (35 items)	Mean	.04	-.04**
	SD	.52	.60**
Science (30 items)	Mean	-.03	.03
	SD	.74	.67
TOTAL	Mean	.01	.02
	SD	.62	.62
Professional Knowledge			
(105 items)	Mean	.03	.06
	SD	.58	.65

* Negative DIF values indicate that male examinees perform better than female examinees with comparable test scores; positive DIF values indicate the reverse.

** One item not scored.

Table 8

Correlations of DIF Index Values on the Same Form
at Two Administrations of the NTE Core Battery

	Subgroups Compared		
Reported Score	Female/Male	Black/White	Hispanic/White
Communication Skills			
Listening (40 items)	.91	.90	.51
Reading (30 items)	.89	.71	.37
Writing (45 items)	.91	.91	.70
General Knowledge			
Social Studies (30 items)	.95	.93	.87
Mathematics (25 items)	.94	.86	.71
Literature & Fine Arts (35 items)	.93	.93	.63
Science (30 items)	.97	.92	.55
Professional Knowledge (105 items)			
Form 1	.93	.81	.52
Form 2	.91	.82	.57

Table 9

Agreement of Flagged NTE Questions on the Same Form at Two Administrations: Black/White Comparison ($|DIF| \geq 1$)

	Number of Items Flagged at Both Administrations	Total Number of Items Flagged	% Agreement of Flagged Items
Communication Skills			
Listening	2	5	40
Reading	1	3	33
Writing	1	6	17
TOTAL	4	14	29
General Knowledge			
Social Studies	6	7	86
Mathematics	0	1	0
Literature & Fine Arts	2	2	100
Science	1	3	33
TOTAL	9	13	69
Professional Knowledge			
Form 1	4	11	36
Form 2	3	10	30

Table 10

Agreement of Flagged NTE Questions on the Same Form at Two Administrations: Hispanic/White Comparison ($|DIF| \geq 1$)

	Number of Items Flagged at Both Administrations	Total Number of Items Flagged	% Agreement of Flagged Items
Communication Skills			
Listening	0	5	0
Reading	0	5	0
Writing	0	4	0
TOTAL	0	14	0
General Knowledge			
Social Studies	2	7	29
Mathematics	1	4	25
Literature & Fine Arts	0	2	0
Science	0	3	0
TOTAL	3	16	19
Professional Knowledge			
Form 1	2	26	8
Form 2	5	21	24

Table 11

Agreement of Flagged NTE Questions on the Same Form
at Two Administrations: Female/Male Comparison
($|DIF| \geq 1$)

	Number of Items Flagged at Both Administrations	Total Number of Items Flagged	% Agreement of Flagged Items
Communication Skills			
Listening	2	6	33
Reading	0	1	0
Writing	2	3	67
TOTAL	4	10	40
General Knowledge			
Social Studies	1	3	33
Mathematics	3	4	75
Literature & Fine Arts	2	3	67
Science	3	5	60
TOTAL	9	15	60
Professional Knowledge			
Form 1	7	15	47
Form 2	8	19	42

Tables 9 through 11 summarize the extent to which the same items are consistently flagged as having absolute values of DIF equal to or greater than one. These tables compare the total number of items flagged in both administrations to the total number of items flagged in at least one administration. As expected, the smaller sample sizes in the Hispanic/White analyses are associated with the lowest levels of agreement between sets of flagged items. The general levels of agreement for the female/male and Black/White samples are similar, despite the larger sample sizes in the female/male analyses. In the female/male comparisons, the agreement ranges from a low of 40 percent for the Communication Skills test to a high of 60 percent for the General Knowledge test. In the Black/White comparisons, the agreement ranges from 29 percent for Communication Skills to 69 percent for General Knowledge. In the Hispanic/White comparisons, the agreement ranges from 0 percent for Communication Skills to 24 percent for one of the two forms of Professional Knowledge.

In the context of this paper, the relatively low percent agreement among flagged items suggests that, in developing hypotheses about causes of DIF, one must examine groups of items as well as individual items since the mean DIF value of the group of items will be more stable than the DIF values of individual items. (Note that the analyses performed to answer Questions 2 and 3 in this report are based on the largest samples. Thus, in the Hispanic/White content analyses, the smallest sample is 415 examinees. Therefore, the results reported for Hispanic/White analyses in the remainder of the paper are probably more stable than in the replication study, but should still be interpreted with caution.)

Question 2: What percentages of the questions are identified as differentially difficult?

Table 12 presents the number and percentage of items flagged in the Black/White analyses of the Communication Skills test, the General Knowledge test, and the Professional Knowledge test. Tables 13 and 14 present the same data for the Hispanic/White and female/male analyses, respectively. Generally, the smallest percentages of items (averaging about 6 percent) are flagged in the Hispanic/White analyses, and slightly larger percentages (averaging about 8 percent) of the items are flagged in the Black/White analyses; in the female/male analyses, about 9 percent of the items are flagged as showing differential difficulty. The flagged items are approximately evenly divided between those on which the reference groups (White or male) perform better than the matched focal groups (Black, Hispanic, or female) and those items for which the reverse is true.

Table 12

Number and Percentage of Items Flagged
in the Black/White Analyses

	3/84				3/85			
	Total Number of Items		DIF* ≤ - 1	DIF** ≥ 1	Total Number of Items		DIF* ≤ - 1	DIF** ≥ 1
Communication Skills								
Listening	40	N	2	3	40	N	1	1
		%	5	8		%	3	3
Reading	30	N	0	1	30	N	1	1
		%	0	3		%	3	3
Writing	45	N	2	3	44	N	1	1
		%	4	7		%	2	2
TOTAL	115	N	4	7	114	N	3	3
		%	3	6		%	3	3
General Knowledge								
Social Studies	29	N	5	3	30	N	4	3
		%	17	10		%	13	10
Mathematics	25	N	2	0	25	N	0	0
		%	8	0		%	0	0
Literature & Fine Arts	35	N	1	2	34	N	1	1
		%	3	6		%	3	3
Science	30	N	2	0	30	N	1	1
		%	7	0		%	3	3
TOTAL	119	N	10	5	119	N	6	5
		%	8	4		%	5	4
Professional Knowledge								
	105	N	2	3	105	N	3	2
		%	2	3		%	3	2

* Negative values indicate that White examinees perform better than Black examinees with comparable test scores.

** Positive values indicate that Black examinees perform better than White examinees with comparable test scores.

Table 13

Number and Percentage of Items Flagged
in the Hispanic/White Analyses

	3/84				3/85			
	Total Number of Items		DIF* ≤ - 1	DIF** ≥ 1	Total Number of Items		DIF* ≤ - 1	DIF** ≥ 1
Communication Skills								
Listening	40	N	0	0	40	N	0	1
		%	0	0		%	0	3
Reading	30	N	0	0	30	N	1	1
		%	0	0		%	3	3
Writing	45	N	2	0	44	N	0	1
		%	4	0		%	0	2
TOTAL	115	N	2	0	114	N	1	3
		%	2	0		%	1	3
General Knowledge								
Social Studies	29	N	0	3	30	N	1	2
		%	0	10		%	3	7
Mathematics	25	N	0	0	25	N	2	0
		%	0	0		%	8	0
Literature & Fine Arts	35	N	0	1	34	N	0	0
		%	0	3		%	0	0
Science	30	N	1	1	30	N	1	0
		%	3	3		%	3	0
TOTAL	119	N	1	5	119	N	4	2
		%	1	4		%	3	2
Professional Knowledge								
	105	N	4	3	105	N	6	5
		%	4	3		%	6	5

* Negative values indicate that White examinees perform better than Hispanic examinees with comparable test scores.

** Positive values indicate the Black examinees perform better than White examinees with comparable test scores.

Table 14

Number and Percentage of Items Flagged
in the Female/Male Analyses

	3/84				3/85			
	Total Number of Items		DIF* ≤ - 1	DIF** ≥ 1	Total Number of Items		DIF* ≤ - 1	DIF ≥ 1
Communication Skills								
Listening	40	N	2	2	40	N	4	1
		%	5	5		%	10	3
Reading	30	N	0	0	30	N	0	1
		%	0	0		%	0	3
Writing	45	N	1	1	44	N	1	2
		%	2	2		%	2	5
TOTAL	115	N	3	3	114	N	5	4
		%	3	3		%	4	4
General Knowledge								
Social Studies	29	N	2	2	30	N	2	1
		%	7	7		%	7	3
Mathematics	25	N	1	1	25	N	2	1
		%	4	4		%	8	4
Literature & Fine Arts	35	N	1	1	34	N	1	2
		%	3	3		%	3	7
Science	30	N	4	4	30	N	2	3
		%	13	13		%	7	10
TOTAL	119	N	8	8	119	N	7	7
		%	7	7		%	6	6
Professional Knowledge								
	105	N	4	3	105	N	4	5
		%	4	3		%	4	4

* Negative values indicate that male examinees perform better than female examinees with comparable test scores; positive values indicate the reverse.

** Positive values indicate that Black examinees perform better than White examinees with comparable test scores.

In the Black/White analyses, a total of 17 items (7 percent) are flagged in the Communication Skills test, with 10 items on which Black test takers perform better than White test takers with comparable test scores and 7 items on which the reverse is true. In the Professional Knowledge test, a total of 10 items (5 percent) are flagged, with 5 items on which Black test takers perform better than White test takers with comparable test scores and 5 items on which the reverse is true. In the General Knowledge test, 26 items (11 percent) are flagged, with 10 items on which Black test takers perform better than White test takers with comparable test scores and 16 items on which the reverse is true.

In the Hispanic/White analyses, a total of 6 items (3 percent) are flagged in the Communication Skills test, with 3 items on which the Hispanic test takers perform better than White test takers with comparable test scores and 3 items on which the reverse is true. In the Professional Knowledge test, a total of 18 items (9 percent) are flagged, with 8 items on which the Hispanic test takers perform better than White test takers with comparable test scores and 10 items on which the reverse is true. In the General Knowledge test, a total of 12 items (5 percent) are flagged, with 7 items on which the Hispanic test takers perform better than White test takers with comparable test scores and 5 items on which the reverse is true.

In the female/male analyses, a total of 15 items (7 percent) are flagged in the Communication Skills test, with 7 items on which the female test takers perform better than the male test takers with comparable test scores and 8 items on which the reverse is true. In the Professional Knowledge test, a total of 16 items (8 percent) are flagged, with 8 items on which the female test takers perform better than the male test takers with comparable test scores and 8 items on which the reverse is true. In the General Knowledge test, a total of 30 items (13 percent) are flagged, with 15 items on which the female test takers perform better than the male test takers with comparable test scores and 15 items on which the reverse is true.

On the average, 6 percent of the items in the Communication Skills test are flagged per group comparison, 7 percent of the items in the Professional Knowledge test are flagged per group comparison, and 10 percent of the items in the General Knowledge test are flagged per group comparison. There is relatively little overlap in the items flagged across all the sets of group comparisons (female/male, Black/White, and Hispanic/White). Among the 137 different items flagged in at least one of the three group comparisons, only 12 items are flagged in two comparisons, and only 1 item is flagged in all three comparisons. Thus, items identified as showing differential item functioning in one of the group comparisons are rarely identified as showing differential performance in the other comparisons. Different items are flagged in the group comparisons in part because of real differences among the groups and in part because of sampling error.

Question 3: What are some of the characteristics of the test questions that are associated with differential difficulty?

Table 15 summarizes the total number of variables studied and the number of variables studied that contain five or more items in all of the categories for the Communication Skills, General Knowledge, and Professional Knowledge tests, respectively. For example, Table 15 indicates that six variables were originally hypothesized as possibly related to differential item functioning. Only three of those variables had at least five items in the categories studied. In cases with very few items in a cell, one item with an extreme DIF value sometimes will result in significant differences among the categories. In this paper we discuss variables with significant differences only when the significant differences appear to be based on data for at least five items in the most extreme cells. With 81 variables and 3 significance tests (1 for each comparison), results for 243 significance tests are available for discussion. At the .05 level, 12 significant values might be found by chance alone. For this reason, our discussion will focus on variables significant at the .01 level, although selected findings at the .05 level will also be discussed.

The full set of significance tables is presented in Appendix B, and significant variables are summarized in Tables 16-18 in the main body of this report. The reader may also want to refer to the detailed descriptions of the variables and classification rules in Appendix A. A summary of results and preliminary observations about the content characteristics that may be related to DIF values are presented below for the Communication Skills, General Knowledge, and Professional Knowledge tests, respectively. Finally, tentative conclusions about variables common to the three tests will be discussed. These observations and tentative conclusions are intended to suggest further study rather than to be definitive. The results must be interpreted cautiously because of the post hoc nature of the study. In an experimental study, the items could have been chosen to allow the variables to be studied independently and to control for many confounding variables. The tests studied here were not constructed for the purpose of experimental study, and, therefore, the variables frequently overlap and the results may be confounded by various other factors. Table 16 summarizes the variables on the Communications Skills test that are significant at the .05 level or beyond.

Table 15

Summary of NTE Variables Studied

Content Area	Number of Variables	Number of Variables With Fewer than 5 Items Per Cell
Communication Skills		
Listening	6	3
Reading	13	7
Writing	9	5
General Knowledge		
Social Studies	9	5
Mathematics	6	4
Literature & Fine Arts	12	10
Science	8	4
Professional Knowledge		
	18	13

Table 16

Summary of Communication Skills Variables
Significant at the .05 Level and Beyond

	Significance Level		
Section/Variable	Female/Male Comparison	Black/White Comparison	Hispanic/White Comparison
Reading			
Set Discrete	NS	NS	.02
Named Gender	.04	NS	.01
Listening			
Item Type	.00	NS	.02
Specifications	NS	NS	.04
Writing			
Item Type	NS	.00	.00

NS = not significant at the .05 level.

Table 17

Summary of General Knowledge Variables
Significant at the .05 Level and Beyond

Section/Variable	Significance Level		
	Female/Male Comparison	Black/White Comparison	Hispanic/White Comparison
Social Studies			
Item Type	NS	.03	NS
Picture/Nonpicture	NS	.02	.04
Content Objective*	NS	.05	NS
Mathematics			
Picture/Nonpicture	NS	.01	NS
Science			
Subject Matter	.05	NS	NS
Specifications*	.00	NS	NS
Literature and Fine Arts			
Discipline	.00	NS	.04
Specifications	NS	.04	.02

* Although data for one or more cells are based on fewer than five items, those cells with the largest absolute values are based on adequate numbers of items.

NS = not significant at the .05 level.

Table 18

Summary of Professional Knowledge Variables
Significant at the .05 Level and Beyond

Section/Variable	Significance Level		
	Female/Male Comparison	Black/White Comparison	Hispanic/White Comparison
Minority Reference Stimulus*	.00	NS	NS
Gender Reference Stimulus*	.00	NS	NS
Negative Stem	NS	.04	NS
Question Format	.03	NS	NS
Specifications	.00	NS	NS
Gerund Response	NS	.03	NS
Question Referent*	.00	NS	NS
Double Factor	.04	NS	NS
Legal Knowledge	.00	NS	NS
Professional Judgment	.05	NS	NS
Controversial Topics	.00	NS	NS

* Although data for one or more cells are based on fewer than five items, those cells with the largest absolute values are based on adequate numbers of items.

NS = not significant at the .05 level.

Few differences in the patterns of performance among the groups are found in the Communication Skills test. In the reading section of the Communication Skills test, only one variable is significant at the .01 level in all the analyses among variables with at least five items in the relevant categories. A significant difference in mean DIF values among the categories in the named gender variable is found in the Hispanic/White analysis. The mean DIF value for items in which both men and women are named is .36. If no one is named or if only a male is named, the mean DIF values are -.10 and -.13, respectively (a description of the categorization of this variable is given on page A11). The data suggest that Hispanic examinees may perform better than a matched group of White examinees on questions in which both men and women are named.

In the listening section of the Communication Skills test, only one finding is significant at the .01 level. A significant difference in the mean DIF values among the categories in the item type variable is found in the female/male comparison. The mean DIF values for the four item types are:

Statements	Questions	Dialogues	Mini-talks
-.08	.34	.36	-.49

These data suggest that women may perform better than a matched group of men on the questions and dialogues item types, and that women may perform less well than comparable men on the mini-talk item type (see page A5 for descriptions of the four item types).

In the writing section of the Communication Skills test, two findings are significant at the .01 level. A significant difference in the mean DIF values between the two categories of the item type variable was found in both the Black/White and Hispanic/White comparisons. The mean DIF values for the two item types are as follows:

	Black/White	Hispanic/White
Usage	.30	.14
Sentence Correction	-.26	-.22

These data suggest that both Black and Hispanic examinees may perform better on usage items and less well on sentence correction items than do a matched group of White examinees (see page A12 for examples of these item types).

Table 17 summarizes the variables that are significant at the .05 level or beyond on the General Knowledge test. In the social studies section of the General Knowledge test, no findings are significant at the .01 level. However, the content objectives variable (alpha = .05) in the Black/White comparison may be of interest (see page A31 for descriptions of the content objectives). The two extreme mean DIF values are for the United States historical category (.56) and the social studies methodology category (-.46). The social studies methodology items typically require the reading of tables, graphs, and

maps and are more quantitative in nature than are the other social studies items. The social studies methodology items also contribute heavily to the differences found between the mean DIF values for items with and without picture stimuli (see page A31 for a description of this variable). Items containing maps, graphs, tables, and charts are classified as picture items and are usually also classified as methodology items. These data suggest that Black examinees may perform better on questions about United States history and less well on questions about social studies methodology than do a matched group of White examinees.

In the mathematics section of the General Knowledge test, only one finding is significant at the .01 level. There is a significant difference in the mean DIF values between the two categories of the picture stimulus variable in the Black/White comparison. The mean DIF value for picture stimulus items is -.25, and the mean DIF value for nonpicture stimulus items is .15 (see page A23 for a description of this variable). These data suggest that Black examinees may perform less well on picture stimulus items than do a matched group of White examinees. Picture stimulus items in mathematics usually contain a table, graph, or geometric figure.

In the science section of the General Knowledge test, only one finding is significant at the .01 level. There is a significant difference in the mean DIF values among the categories of the specifications variable (see pages A25-A28) in the female/male comparison. The two categories with the most extreme DIF values are living things (.47) and matter (-.60). This is consistent with the finding (alpha = .05) for the subject matter variable (the subject matter variable is described on page A29). The mean DIF values for the categories in the subject matter variable are as follows:

	Female/Male
Biology	.28
Chemistry	-.35
Physics	-.44
Other	-.15

These data suggest that women may perform better than a matched group of men on biology items, and that women may perform less well than a matched group of men on chemistry and physics items. This is consistent with the observation that men are more likely to have had course work in chemistry and physics, and that women's courses in science (less numerous than men's) are likely to have been in the biological sciences.

In the literature and fine arts section of the General Knowledge test, only one finding is significant at the .01 level. A significant difference in the mean DIF values among the categories in the discipline variable was found in the female/male comparison. The two categories with the most extreme mean DIF values are architecture (-.59) and performing arts (.66). The mean DIF values for the other categories in the variable--fiction, painting, poetry, and sculpture--are all close to zero (see page A17 for a description of the variable). These data

suggest that men may perform better than a matched group of women on questions about architecture and that women may perform better than a matched group of men on questions about the performing arts.

The specifications variable (see pages A15-A16), which reached the .05 level of significance in both the Black/White and Hispanic/White comparisons, may also be of interest. The mean DIF values for the categories of this variable are similar for the two analyses:

	Black/White	Hispanic/White
Ability to recognize basic elements...	-.19	-.04
Ability to analyze and interpret...	-.06	-.07
Ability to relate works... to their social/historical context	.55	.40

These data suggest that Black and Hispanic examinees may perform better on questions relating to the social/historical contexts of works of art than do a matched group of White examinees.

Table 18 summarizes the variables that are significant at the .05 level or beyond on the Professional Knowledge test. Six variables are significant at the .01 level for the female/male analyses, while no variables are significant at the .01 level for the Black/White analyses or the Hispanic/White analyses.

Significant differences in the mean DIF values among the categories in the minority reference in stimulus and gender reference in stimulus variables (described on page A4) were found in the female/male analyses. Few items on this test refer to a specific minority group or exclusively to men or women. Within both variables, one category was that the stimulus does not refer to people. The mean DIF value for the 33 items that do not refer to people is -.39. Thus, the data suggest that female examinees may do less well than a matched group of male examinees on items that do not refer to people. There are not sufficient data to say whether use of exclusively female referents, male referents, or minority referents in the stimulus would have any impact on differential item difficulty on this test.

The content specifications of the Professional Knowledge test divide the test questions into three categories--items relating only to the process of education, items relating only to the context of education, and items relating to both the process and the context of education. Significant differences among the mean DIF values for items in these three categories were found in the female/male analyses, with mean DIF values of .21, -.33, and .06 for the three categories, respectively. Women performed better than a matched group of male examinees on questions relating to the process of education and less

well than a matched group of male examinees on questions related to the context of education (this variable is described on pages A33-A36).

Significant differences at the .01 level among the mean DIF values in the question referent variable (see pages A33-34) were also found. The categories in this variable indicate whether the question refers to the teacher only; teacher and student; teacher and parent; teacher, student, and parent; student only; student and parent; parent only; or to neither teacher, student, or parent. The mean DIF value of items that refer to teacher, student, and parent is .50, and the mean DIF value for items that refer to none of these is -.32. Thirty-three of the 48 items that do not refer to teacher, student, or parent do not refer to people at all, and thus this finding is interrelated to the results of the minority reference and gender reference variables. The results suggest that female examinees may perform better than a matched group of male examinees on items that mention a parent, student, and teacher.

There were also significant differences in the mean DIF values of the categories in the legal knowledge (see page A34) and controversial topics (see page A35) variables in the female/male analyses. Women found the 14 questions about the legal context of education (e.g., student rights, handicapped students, sex and racial equity, bilingual education, school district organization, federal and state involvement in the governance of education, etc.) more difficult than did a matched group of male examinees (mean DIF = -.64). Women also performed less well than a matched group of male examinees on the 22 questions (mean DIF = -.38) about controversial topics in education (e.g., sex education, evolution, religion, politics). Of the 14 items categorized as requiring legal knowledge, 10 also are categorized as testing the context of education, and the other 4 are categorized as testing both the process and the context of education. Seven of the 14 legal knowledge questions are also categorized as testing controversial topics. Of the 22 items categorized as testing controversial topics, 15 had negative DIF values, and 11 of these 15 were also categorized as measuring the context of education. Thus, these three variables--the specification content areas (process of teaching, context of teaching, or both), legal knowledge, and controversial topics--are somewhat confounded. It would require much larger samples of items in each of these categories to determine specifically how each category contributes to the DIF results.

Although the professional judgment variable is significant only at the .05 level, it is worth noting that there also is confounding of the categories in this variable with those in the legal knowledge, controversial topics, and content category variables.

In addition to the significant findings, several nonsignificant results are worth noting. One of these results relates to the common assertion that questions with a negative stem (e.g., "all of the following EXCEPT," "which of the following is NOT") will be differentially difficult for some groups. This hypothesis could not be tested for Communication Skills, since only one negative stem item appears in the studied tests. In the General Knowledge test, no

significant differences in the mean DIF values of negative and positive stem items are found for the Black/White, Hispanic/White, or female/male comparisons. In the Professional Knowledge test, the Hispanic/White and female/male mean DIF values in the two categories are also not significantly different. However, the negative stem items are slightly more difficult for Black examinees than for a matched group of White examinees. (The mean DIF value of the 23 negative stem items is -.20, while the mean DIF value of the 152 positive stem items is .02, alpha = .04.)

A second common hypothesis is that Roman numeral format items (items that present three or four responses identified by Roman numerals and ask the examinee to choose among options that present various combinations of the Roman numbers (e.g., I and II only, I, II, and III, etc.) will be differentially difficult for some subgroups. No significant differences between Roman numeral format and other formats are identified in any of the tests.

Conclusions

The conclusions are organized by the three questions that form the framework for the study.

Question 1: Are the same questions identified as showing differential item functioning in analyses of test takers from two different administrations of the same test?

The correlations of DIF index values from two administrations of the same test range from .89 to .97 in the female/male analyses, from .71 to .93 in the Black/White analyses, and from .37 to .87 in the Hispanic/White analyses. The correlation analysis gives a somewhat more positive picture of the replicability of the results than does the analysis of the consistency with which the same items are identified as having absolute values of DIF equal to or greater than one.

In the test development process, we are dedicated to eliminating items that show extreme DIF values for reasons that appear to be related to group membership rather than to the construct being measured. One prerequisite for accomplishing this is the reliable identification of items with such DIF values.

In the analysis of the consistency of the identification of items with absolute values of DIF equal to or greater than one, there is 40 to 60 percent agreement in the female/male analyses, 29 to 69 percent agreement in the Black/White analyses, and 0 to 24 percent agreement in the Hispanic/White analyses. The lower replicability of the Hispanic/White analyses may well be due to lower sample sizes. In view of the large sample sizes of the female/male and Black/White analyses, this is modest replicability.

Question 2: What percentages of questions are identified as differentially difficult?

Generally, about 6 percent of the items in the Hispanic/White analyses and about 8 percent of the items in the Black/White analyses were identified as differentially difficult. In the male/female analyses, about 9 percent of the items show differential difficulty. The flagged items are approximately evenly divided between those on which the reference groups perform better and those on which the focal groups perform better. The finding that a greater number of items are flagged in the female/male analyses than in the Black/White and Hispanic/White analyses is somewhat surprising. Because smaller sample sizes tend to yield larger DIF index values (Wright, 1986), the finding is not explained by the difference in sample size. It appears that, after matching groups on the criterion score, there are more differences remaining between the female and male groups than there are between the Black and White groups and between the Hispanic and White groups. It is interesting to note that relatively few items are flagged in more than one of the three analyses.

Combining the data across all three group comparisons in order to obtain average flagging rates for the three tests, 6 percent of the items in the Communication Skills test are flagged, 7 percent of the items in the Professional Knowledge test are flagged, and 10 percent of the items in the General Knowledge test are flagged. The relatively lower proportion of flagged items in the Communication Skills test is consistent with our previous findings of relatively low proportions of flagged items in verbal sections of two other tests, the Graduate Management Admission Test and the Graduate Record Examinations General Test (Wild & McPeek, 1986), and with findings for the Scholastic Aptitude Test (Dorans & Kulick, 1986). Perhaps the different patterns of course-taking and the different interests of the groups are more likely to lead to differential item functioning on Professional and General Knowledge rather than Communication Skills.

Question 3: What are some of the characteristics of test questions that are associated with differential difficulty?

To evaluate the content characteristics of questions identified as differentially difficult, 81 variables were hypothesized and tested for each of three comparisons (Black/White, Hispanic/White, female/male). Of the 243 significance tests, only 13 were significant at the .01 level or beyond and had at least five items in each cell. Twelve of the 81 variables are represented in the 13 significant F-tests. Of the 13, 2 were found in Black/White comparisons, 2 were found in Hispanic/White comparisons, and 9 were found in female/male comparisons. These differences appear to be primarily related to specific areas of academic knowledge rather than to format characteristics of the items.

Few significant differences are found in the Communications Skills test. The most notable are the differences among item types in the listening section for female and male test takers, and the differences between item types in the writing section for Black and White, and Hispanic and White, test takers. The different item types in listening and writing measure different knowledges and skills.

There are no significant differences in the social studies section of the General Knowledge test at the .01 level. However, the data do suggest that Black examinees may perform better on questions about United States history and less well on questions about social studies methodology than do a matched group of White examinees. In the mathematics section of the General Knowledge test, Black examinees appear to perform less well on picture stimulus items than do a matched group of White examinees. Both picture stimulus items in mathematics and social studies methodology items are usually based on information contained in tables, graphs, or figures. The data from the science section of the General Knowledge test suggest that women may perform better than a matched group of men on biology items, and that women may perform less well than a matched group of men on chemistry and physics items. The data from the literature and fine arts section of the General Knowledge test suggest that men may perform better than a matched group of women on questions about architecture, and that women may perform better than a matched group of men on questions about the performing arts. There is also some indication that Black and Hispanic examinees may perform better than a matched group of White examinees on questions relating to the social/historical contexts of works of art.

Significant differences on the Professional Knowledge test are found primarily in the female/male analyses on variables relating to test content. The data suggest that women may perform better than a matched group of men on questions relating to the process of education and less well than a matched group of men on questions related to the context of education. The data suggest that women may perform better than a matched group of men on items that mention a parent, student, and teacher. Women appear to perform less well than a matched group of men on questions relating to the legal context of education, and on questions about controversial topics in education. It would be interesting in the future to investigate whether these findings are related to the different course-taking patterns of men and women.

Little support was found in these data for the hypotheses that negative stem items and Roman numeral format items would be differentially difficult for some subgroups.

Overall, relatively few significant differences were found, and these tend to be in specific areas of knowledge. It is quite likely that many of these differences in knowledge are related to the particular undergraduate courses taken by the different groups. For example, women take more courses than do men in elementary education, and men take more courses than do women in school administration. These differences in undergraduate training are probably reflected in the finding that women perform better than a matched group of men on questions about the process of education, including pedagogy. They are

probably also reflected in the finding that men perform better than a matched group of women on questions relating to the legal context of education and to controversial topics in education. Future research should investigate differential item functioning within groups of test takers with similar training and preparation to examine this hypothesis about the findings.

In addition, the finding that differential item functioning appears to be related to specific content knowledge emphasizes the importance of having detailed content specifications for the test, such as those presented in A Guide to the NTE Core Battery Tests. These findings about the performance of different groups of test takers will provide valuable information to the committees of teachers and teacher educators as they evaluate the content specifications for new forms of the tests.

Additional data are required to test many of the hypotheses generated for this study about the possible correlates of differential difficulty. A number of additional test forms will need to be analyzed in order to have enough questions to assess the test content specifications' relationship to differential item difficulty, because there are typically only one or two items per test form testing a particular content area. Once hypotheses that may account for differential item difficulty have been identified, research studies to confirm or disconfirm the hypotheses will need to be conducted. The experimental manipulation of test questions especially designed to test particular hypotheses will be required. The implications of these findings (both retrospective and experimental) would need to be carefully assessed by both internal and external advisory committees in order to make decisions about the implications for test content.

References

Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 10, 95-106.

Cardall, C., & Coffman, W. E. (1964). A method for comparing the performance of different groups on the items in a test (Research Bulletin RB-64-61). Princeton, NJ: Educational Testing Service.

Carlton, S. T., & Harris, A. M. (1989a). Female-male performance difference on SAT: Causes and correlates. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Carlton, S. T., & Harris, A. M. (1989b). Characteristics associated with differential item performance on the SAT: Selected ethnic group comparisons. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco.

DeMauro, G., & Huffman, R. (1989). Ethnic group performance on a listening skills test. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco.

Dorans, N. J., & Kulick, E. M. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. Journal of Educational Measurement, 4, 355-368.

Dwyer, C. A. (1986). Using indices of differential item functioning in test development decisions. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Holland, P. W. (1985). On the study of differential item performance without IRT. Proceedings of the Military Testing Association.

Holland, P. W., & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.

Marascuilo, L. A., & Slaughter, R. E. (1981). Statistical procedures for identifying possible sources of item bias using the X^2 statistics. Journal of Educational Measurement, 18, 229-248.

McPeek, W. M., & Wild, C. L. (1986). Performance of the Mantel-Haenszel statistic in a variety of situations. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

McPeek, W. M., & Wild, C. L. (1987). Characteristics of quantitative items that function differently for men and women. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Mellenberg, G. J. (1982). Contingency table models for assessing item bias. Journal of Educational Statistics, 7, 105-118.

O'Neill, K. A., Wild, C. L., & McPeek, W. M. (1989). Gender-related differential items performance on graduate admissions tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

O'Neill, K. A., Wild, C. L., & McPeek, W. M. (1989). Characteristics of GRE verbal test items that show differential item functioning for Black and White examinees. Paper presented at the National Council of Measurement in Education, San Francisco, CA.

Phillips, A., & Holland, P. W. (1986). A new estimator of the variance of the Mantel-Haenszel log-odds-ratio estimator (Research report no. 86-21). Princeton, NJ: Educational Testing Service.

Scheuneman, J. D. (1979). A method of assessing bias in test items. Journal of Educational Measurement, 16, 143-152.

Wild, C. L., & McPeek, W. M. (1986). Performance of the Mantel-Haenszel statistic in identifying differentially functioning items. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Wright, D. (1986). An empirical comparison of the Mantel-Haenszel and standardization methods of detecting differential item performance. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Zieky, M. (1988). Differential item difficulty in test development at Educational Testing Service. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

APPENDIX A

Item Classifications

This appendix presents a detailed description of the classification scheme used to investigate NTE tests in this research study. These classifications are used to investigate whether specific item characteristics are related to differential item functioning values.

The chart on the following page summarizes the variables that were used by test developers to categorize items in the following tests:

Communication Skills:	Listening
	Reading
	Writing
General Knowledge:	Literature and Fine Arts
	Mathematics
	Science
	Social Studies
Professional Knowledge	

It must be noted that test developers did, in some instances, use the same generic titles as classification variables for different tests, but with different divisions and meanings. Such classification variables have the word varies in parentheses next to them. For example, the classification variable Specifications I (varies) has different divisions and different meanings for each of the tests studied. In contrast, four variables--minority stimulus, gender reference in the stimulus, negative item, and Roman numeral format--are the same for all tests to which they apply. Following the chart are the descriptions of how items were classified within each variable for each test. These descriptions will clarify the differences in classification schemes among tests.

Summary of Variables on Which Items Were Classified

	Communication Skills			General Knowledge				
Variables	Listening	Reading	Writing	Lit. & Fine Arts	Math.	Science	Social Studies	Professional Knowledge
Minority Stimulus	X	X	X	X	X	X	X	X
Gender Reference in Stimulus	X	X	X	X	X	X	X	X
Negative Item	NA	X	NA	X	X	X	X	X
Roman Numeral Format	NA	X	NA	X	NA	X	X	X
Specifications I (varies)	X	X	X	X	X	X	X	X
Specifications II (varies)	X	X	X	X	X	X	X	X
Speaker Gender	X							
Item Type (varies)	X		X					
Content (varies)		X	X		X			
Subject Matter		X				X		
Passage Length		X						
Sets Versus Discretes (varies)		X		X			X	
Question Format (varies)		X						X
Named I--Minority		X						
Named II--Gender		X						
Underline Position			X					
Key			X					
Picture Stimulus (varies)				X	X	X	X	X
Skills (varies)				X		X	X	
Content Objectives							X	
Single Versus Multiple Stimuli				X				

Summary of Variables on Which Items Were Classified (continued)

	Communication Skills			General Knowledge				
Variables	Listening	Reading	Writing	Lit. & Fine Arts	Math.	Science	Social Studies	Professional Knowledge
Discipline				X				
Culture				X				
Time Period				X				
Response Length								X
Gerund Responses								X
Evaluation of Action								X
Question Referent								X
Options Are Verbal Responses								X
Double Factor								X
Ordering Format								X
Legal								X
Judgment/Consensus								X
Complexity								X
Controversial Topics								X
Theorists/Theories								X

VARIABLES COMMON TO ALL TESTS

MINORITY STIMULUS

1 = Black	Stimulus refers to Black people.
2= Hispanic	Stimulus refers to Hispanic people.
3 = Other Minority	Stimulus refers to other minorities.
4 = General	Stimulus refers to people of no specified ethnic origin.
5 = Nothing	Stimulus does not refer to people.

GENDER REFERENCE IN STIMULUS

1 = Female	Stimulus refers to females.
2 = Male	Stimulus refers to males.
3 = Mixed or Unidentifiable	Stimulus refers to both males and females.
4 = Neutral	Stimulus does not refer to people.

NEGATIVE ITEM (No/Except)	(Does not apply to Communication Skills listening or writing)
1 = Negative Stem	Use "NOT," "CANNOT," "EXCEPT," "LEAST," "INCORRECT," "FALSE," etc., in stem
2 = Positive Stem	Do not use "NOT," "CANNOT," etc., in stem

ROMAN NUMERAL FORMAT	(Does not apply to Communication Skills listening or writing, or General Knowledge math)
1 = Roman	Involves Roman numeral format
2 = Non-Roman	Does not involve Roman numeral format

VARIABLES STUDIED IN COMMUNICATION SKILLS: LISTENING

(See also minority stimulus and gender reference in stimulus in the section "Variables Common to All Tests.")

ITEM TYPE

1 = Statements	Candidates hear a short statement; the answer choices consist of short statements, that are either paraphrases of the original statement or direct inferences from the original statement. This item type has no stem.
2 = Questions	Candidates hear a question; they are asked to pick an appropriate response to the question.
3 = Dialogues	Candidates hear a dialogue, about which they answer questions.
4 = Mini-talks	Candidates hear a short talk, about which they answer questions.

SPECIFICATIONS

I = Basic Comprehension

This category of items tests the candidates' understanding of the spoken message they have just heard. Items in this category do not require inference or evaluation. It is further divided into the following subskills:

11 = IA	Noting details
12 = IB	Paraphrasing
13 = IC	Recognizing cognitive clues, e.g., inferring the meaning of a word from its context
14 = ID	Understanding connotations of words
15 = IE	Interpreting figurative language
16 = IF	Identifying the main idea

SPECIFICATIONS (continued)

II = Analysis

Items in this category test the candidates' ability to analyze and interpret what they have heard. It is divided into the following subskills:

21 = IIA	Identifying assumptions made by the speaker
22 = IIB	Drawing inferences from the information contained in the message
23 = IIC	Recognizing the implications of assertions made in the message, e.g., causation, possible consequences of an action
24 = IID	Identifying the speaker's tone

Statements (item type 1 above) that require the candidate to make an inference based on the statement or to recognize the implications of a statement are classified IIB or IIC.

III = Evaluation

Items in this category test the candidates' ability to identify and evaluate the logical structure of the message they have heard. This category is divided into the following subskills:

31 = IIIA	Identifying and evaluating how a persuasive argument is constructed
32 = IIIB	Assessing the appropriateness of supporting material
33 = IIIC	Evaluating the effect of the speaker's tone on the audience

IV = Feedback

Items in this category test the candidates' ability to identify appropriate responses to questions. Only the item type "questions" is placed in this category.

41 = IVA	Feedback - Response

SPECIFICATIONS (continued)

SPEAKER GENDER (Stimulus)

1 = Female Speaker

2 = Male Speaker

3 = Mixed

The people who read the test on the tape are male and/or female. This category refers to the gender of the speaker(s) only, not to the content of the test.

VARIABLES STUDIED IN COMMUNICATION SKILLS: READING

(See also minority stimulus, gender reference in stimulus, negative item, and Roman numeral format in the section "Variables Common to All Tests.")

CONTENT

1 = Teacher Related	All passages or statements that deal with matters related to education (classroom issues, educational policy, history of education, etc.) are classified as 1; all others are classified as 2.
2 = General Interest	

SUBJECT MATTER

1 = Science/Technology	All passages or statements that deal with science or technology or specialists in these fields are classified as 1; all others are classified as 2. Category 1 should be broadly interpreted to include all topics in which scientific or technological development is a key feature.
2 = Other	

PASSAGE LENGTH

1 = Long Passage	Long passages contain approximately 200 words.
2 = Short Passage	Short passages contain approximately 100 words.
3 = Statements	Statements are generally one sentence (sometimes two) containing fewer than 100 words.

SETS VERSUS DISCRETES

1 = Set	Sets contain more than one question based on a passage or statement.
2 = Discrete	A discrete question is a single question based on a passage or statement.

QUESTION FORMAT

1 = Closed	A closed question is one in which the actual question being asked is a complete sentence and the answer choices are independent units (words, phrases, or sentences).
2 = Open	An open question is one in which the actual question being asked is not a complete sentence and is completed by the answer choices.

SPECIFICATIONS I

COMPREHENSION

11 = Main idea--This kind of question tests the ability to grasp the main idea of a passage.

12 = Supporting idea/detail--This kind of question tests the ability to grasp secondary or supporting ideas, that is, ideas or details provided in support of the author's main idea.

13 = Relationships--This kind of question tests the ability to grasp relationships among ideas in the passage, e.g., relationships of cause and effect, and sequence.

14 = Paraphrase/summary--This kind of question tests the ability to identify accurate paraphrases or summaries of elements in the passage.

ANALYSIS

21 = Author's purpose--This kind of question tests the ability to discern the purpose or intent of the author in writing the passage.

22 = Author's assumptions--This kind of question tests the ability to discern assumptions upon which the author bases the argument of the passage.

23 = Author's attitude or tone--This kind of question tests the ability to recognize the expression of an attitude of the author toward persons or ideas, or the expression of a particular tone in the writing of the passage.

24 = Implications/inferences--This kind of question tests the ability to perceive what is implied or suggested rather than directly stated by the author and to make inferences from the directly stated content of the passage.

SPECIFICATIONS (continued)

25 = Fact versus opinion--This kind of question tests the ability to recognize the difference between objectively stated fact and subjectively stated opinion.

26 = Organization--This kind of question tests the ability to understand how the passage is organized and how that organization contributes to the effectiveness of the author's argument.

27 = Use of language--This kind of question tests the ability to understand when and how the author uses language to particular effect or with particular intent.

28 = Application--This kind of question tests the ability to recognize ideas or situations that are analogous to or extensions of ideas or situations presented in the passage.

EVALUATION

31 = Emotional/manipulative aspects of the passage--This kind of question tests the ability to recognize elements in the passage that are intended to arouse the emotions of the reader or to manipulate his or her thinking in a particular direction.

32 = Strengths/weaknesses of the argument--This kind of question tests the ability to identify particular strengths or weaknesses in the author's principal argument or other arguments in the passage.

33 = Relevance/appropriateness of supporting arguments or evidence--This kind of question tests the ability to evaluate facts or ideas, either stated in the passage or extraneous to the passage, with regard to their relevance or appropriateness to the principal argument or other arguments in the passage.

34 = Relation of the passage to the audience--This kind of question tests the ability to make judgments about the nature of the readership envisioned by the author in writing the passage.

SPECIFICATIONS II

10 = Comprehension — Comprehension questions test literal reading skills, that is, the ability to understand what is directly stated in a passage.

SPECIFICATIONS II (continued)

20 = Analysis — Analysis questions test the ability to understand the implied content of a passage, the use of language in the passage, and the organization of the passage.

30 = Evaluation — Evaluation questions test the ability to make critical judgments about a passage.

NAMED I - MINORITY

0 = No One Named

1 = Black American Named

2 = Hispanic American Named

3 = Native American Named

4 = Asian American Named

5 = Other Minority Named

6 = Nonminority Named

NAMED II - GENDER

0 = No One Named

1 = Female Named

2 = Male Named

3 = Both male and female named or person of unknown gender named

VARIABLES STUDIED IN COMMUNICATION SKILLS: WRITING

(See also minority stimulus and gender reference in stimulus in the section "Variables Common to All Tests.")

ITEM TYPE

1 = Usage Item Type. Example:

He spoke bluntly and angrily to we spectators. No error.
A B C D E

2 = Sentence Correction Item Type. Example:

Modern travelers seem to prefer speed to comfort.

(A) to comfort
(B) than comfort
(C) rather than being comfortable
(D) instead of being comfortable
(E) more than comfort

SPECIFICATIONS

Each item containing an error was classified into at least one of the categories listed below. In addition, some of the items were doubly classified when the error encompassed two of the problems.

10-15 REFER TO THE PART OF SPEECH THAT CONTAINS THE ERROR IN THE SENTENCE

10 = Preposition Problem	Misuse of words such as like, as, instead of, for
11 = Noun Problem	Errors of noun number or noun case
12 = Pronoun Problem	Errors of pronoun case (us, we) or number (it, them)
13 = Verb Problem	Errors relating to tense (leading, to lend) or number (is, are)
14 = Adjective Problem	Errors when adjective used for adverb or mismatched with the noun
15 = Adverb Problem	Errors when adverb used as adjective

SPECIFICATIONS (continued)

16-22 REFER TO STRUCTURAL ELEMENTS THAT MAY BE FAULTY

16 = Coordination Problem	Errors in the use of coordinators such as or, nor; and, but
17 = Subordination Problem	Errors in structure such as: A useful way of escaping the notice of their enemies, insects
18 = Correlation Problem	Errors in the use of connectors such as neither, nor; either, or
19 = Comparison Problem	Errors in the use of words such as less, lesser; better, best
20 = Parallelism Problem	Errors in structure such as: The editorial made fun of staff, contained obscenities, and not taking school seriously.
21 = Negation Problem	Misuse of words or phrases such as whether or not
22 = Modification Problem	Errors in structure such as: comments on racial problems in his work humorously
23 = Diction and Word Choice	Includes wrong word choices: allusion for illusion, for example, would be included here
24 = Idiom	Refers to expressions that are simply not idiomatic
25 = Redundancy	Includes expressions such as annually each year
26 = Punctuation and Capitalization	Includes commas, semicolons, and capitalization as well as such errors as it's for its
27 = No Error	

CONTENT OF STIMULUS

1 = Humanities	Humanities includes the arts and human activities not covered by science.

BEST COPY AVAILABLE

CONTENT OF STIMULUS (continued)

2 = Science	Science includes much more than simply biology, chemistry, or physics. Mention of animals, veterinarians, medication, illness, environmental concerns, etc., are included in the science classification.

UNDERLINE POSITION (Sentence Correction Items)

1 = Beginning of Sentence	is any underline that includes the first word of the sentence
2 = Middle of Sentence	is any underline that does not include the first or last word of a sentence
3 = End of Sentence	is any underline that includes the last word of a sentence
4 = Entire Sentence	is any underline that includes the first and last words of a sentence
5 = N/A	

KEY

1 = Error	
2 = No Error	The A key in sentence correction and the E key in usage items are the No Error options.

VARIABLES STUDIED IN GENERAL KNOWLEDGE: LITERATURE AND FINE ARTS

(See also minority stimulus, gender reference in stimulus, negative item, and Roman numeral format in the section "Variables Common to All Tests.")

SETS VERSUS DISCRETES

1 = Linked Sets	A set is made up of one or more stimuli and two or more items.
2 = Discrete	A discrete is a single item based on a stimulus.

SINGLE VERSUS MULTIPLE STIMULI

1 = Single	A single stimulus may be followed by a single item or by a set of items. For example, a passage from a literary work or a photograph of a single artwork may be the focus of one or more items.
2 = Multiple	Two or more stimuli may be followed by a single item or by a set of items. For example, photographs of two buildings may be the focus of one or more items. These items may ask the candidate to compare and contrast aspects of the stimuli presented or the candidate may be asked to identify a style, technique, etc.

PICTURE STIMULUS

1 = Picture	Stimulus contains a picture.
2 = Nonpicture	Stimulus does not contain a picture.

SPECIFICATIONS

1 = I	The ability to recognize basic elements and components of works of literature and fine arts: vocabulary in context, basic musical concepts, widely used symbols, representational features, nature of materials, compositional elements, etc.

SPECIFICATIONS (continued)

2 = II	The ability to analyze and interpret works of literature and fine art by dealing with the interrelationship of the elements of works, including such concepts as relation of parts of works to the whole, pattern, unity, function; and by dealing with such topics as meaning, symbolism, convention, holistic significance, theme, mood, and tone
3 = III	The ability to relate works of literature and fine art to one another and to their social/historical contexts by seeing works as reflections of their social/historical contexts or as embodiments of significant questions concerning human values; by demonstrating awareness of similarities and differences between works from different cultural traditions and historical periods; and by distinguishing between works from different genres, stylistic types, and schools

SKILLS

These skills are an elaboration and combination of the skills outlined in the previous group, labeled "Specifications." The skills identified in the specifications are individual skills, whereas the ones included in this group are a combination and interaction of two or more skills.

1 = Identification	Identification of the relationship that exists between one or more elements (described in I above)--for example, how the use of lighting in a painting affects the mood
2 = Analysis and Interpretation	Use of analysis and interpretation skills to identify either the social/historical context of a work or the social significance of the material presented in the work
3 = Comparison	Use of comparison skills to carry out analysis and interpretation of materials and to place material in its historical context

DISCIPLINE

1 = Fiction	Excerpts from novels and essays
2 = Poetry	Whole poems or excerpts from poems
3 = Painting	Black and white photographic reproductions of entire paintings or details
4 = Sculpture	Black and white photographic reproductions of entire sculptures or details
5 = Architecture	Black and white photographic reproductions of entire buildings or details
6 = Performing Arts	Black and white photographic stills of theater, film, and dance productions

CULTURE

1 = American (White)	Refers to artworks produced by White American cultural groups
2 = European	Refers to artworks produced by European cultural groups
3 = American (Other)	Refers to artworks produced by groups such as Native Americans, Black Americans, Asian Americans, Hispanic Americans, etc.

TIME PERIOD

1 = Pre-1800	Items about works created before 1800
2 = 1800	Items about works created in the 19th century
3 = 1900	Items about works created in the 20th century

VARIABLES STUDIED IN GENERAL KNOWLEDGE: MATHEMATICS

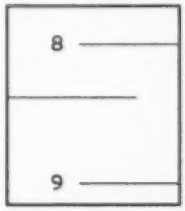
(See also minority stimulus, gender reference in stimulus, and negative item in the section "Variables Common to All Tests.")

ITEM CONTENT

Mathematics items may deal with spatial knowledge, inference, estimation, some combination of these three, or none of these three.

10 = Spatial

Items included in the spatial category are those in which the examinee must make some judgment about the relative size, arrangement, or appearance, after rearrangement, of some figure(s). In the figure below, for example, it is necessary to judge the relative position of the arrow in order to estimate the reading on the scale:



Thus, a test item on reading a graph might be categorized as spatial if it involved judgment of this sort. (If, however, all the bars on a bar graph terminated at calibrated marks that were identified on the scale, the item would be categorized as spatial.) An item was categorized as spatial not because of a content being tested, but because of a skill required for answering the question: a geometry item, therefore, might or might not be classified as spatial.

20 = Inference

A test item categorized as inference is one in which the examinee must infer information needed to solve a problem from one or more pieces of given information. For example, if it is given in the item that 35

ITEM CONTENT Inference (continued)

	percent of the employees of a company take their vacations in August, but the question relates to employees who do not take their vacations in August, the examinee must determine one of the numbers to use in computation, and do this by inference. The examinee is not asked to check the validity of the inference; rather, he or she must determine that an inference is necessary, make it, and then use it to solve a problem.
30 = Estimation	Items categorized as estimation do not require exact answers. Rather, the examinee may be asked to select, from a given range of values, the one that is "close to" or "about" the correct value. An item that asks for the greatest number of nontaxable articles costing $1.95 each that could be purchased for $20 would also be considered an estimation item, since it could be solved by estimation.
12 = Spatial/Inference	Some items require spatial knowledge and inference or estimation. If, for example, the information needed to make an inference or an estimate required a spatial judgment, the item was so classified.
40 = None of the above	

SPECIFICATIONS

We believe these competencies are developmental in nature, accruing not by an individual having taken one or more specific courses, but rather through the cumulative effect of the total mathematics curriculum. There is, of course, an assumption here that certain commonalities exist in all mathematics curricula, and that all NTE candidates will have had certain "basics"; for example, it is assumed that all candidates will have studied fractions, regardless of the city or state in which they went to school, or the textbooks or methodologies used in the schools. We therefore consider these competencies to be neither content specific nor content free.

SPECIFICATIONS (continued)

Each item was coded for the major category (A through F below) as well as for the subskills listed under each category.

1 = A. Has good number sense, that is, understands how numbers behave.

1. Has a sense of order among numbers--e.g., knows that 1/2 is between 1/3 and 2/3, that $-3 < -2$, that 1.9 is closer to 2 than 2-1/4 is, that 75% is less than 1, and that 600% is more than 1.

2. Has a meaningful understanding of the way numbers are named (i.e., place value); understands that a number has many names; and has facility in translating from one to another as needed; e.g., can use 50% or .5 or 1/2--whichever simplifies computation or aids flexibility of thinking.

3. Has a sense of the order of magnitude of a number as it relates to place value or scientific notation; e.g., recognizes that 100 is 1,000 times as great as 0.1, or that 2.57 x 10 is 1/100 as great as 2.57 x 10.

4. Estimates, or otherwise predicts, the outcome of computation.

2 = B. Understands and uses numbers in an appropriate way to quantify thinking.

1. Recognizes an appropriate match between mathematics and real life; e.g., can establish a correct ratio or percentage, or select an appropriate operation for a real-life problem.

2. Recognizes necessary and sufficient conditions for the solution of real-life problems; e.g., for a real-life problem, knows what numbers are needed and how to obtain them (e.g., what measurements are needed).

3. Solves real-life problems by estimating answers and doing the necessary computation.

4. Recognizes multiple ways to find answers (and equivalent computational procedures) and chooses appropriately.

5. Recognizes an appropriate number that can be used as an answer to a problem and adjusts and interprets answers to fit the content of the problem; e.g., the answer 5-1/3 would be recorded as 6 if it represented the number of cars needed to transport people, but recorded as 5 if it represented the number of passengers per car.

SPECIFICATIONS (continued)

6. Correctly predicts the outcome of changing some number or condition in a problem in an "if-then" sense--e.g., if N+5 = Q, what is the value of N+10?

7. Interprets numbers when used to express probability.

3 = C. Recognizes and uses mathematical relationships.

1. Distinguishes among direct, inverse, and other kinds of variations without necessarily knowing the correct term for the relationship.

2. Recognizes spatial relationships in everyday life--e.g., identifies and predicts possible relationships among lines in space.

3. States and uses relationships for the measures of common two- and three-dimensional geometric figures.

4. Symbolizes a relationship appropriately or, conversely, interprets a relationship expressed in symbols; understands the use of a formula as a way to solve a class of similar problems.

5. Recognizes equivalent relationships having a different form--e.g., d = rt <---> t = d - r.

6. Solves problems involving ratio and proportion and percentage.

7. Recognizes relationships evident in data and makes appropriate predications and/or extrapolations from those data.

4 = D. Understands the mathematical basics of measurement.

1. Understands that numbers are used to quantify attributes (e.g., length, temperature, area) of objects, not the objects themselves.

2. Recognizes and uses appropriate units for making everyday measurements.

3. Recognizes and uses geometric concepts in making linear, area, and volume measurements.

4. Understands the relationship between the size of the unit and the number of units--e.g., the shorter the unit used to measure length, the larger the number of those units in a specific measurement.

SPECIFICATIONS (continued)

5. Knows in a general way how to convert from one unit to another in the same system--e.g., whether to multiply or divide.

6. Determines the measurements needed to solve a problem; can solve measurement problems.

7. Is literate about the metric (SI) system.

8. Reads a calibrated scale correctly whether the calibration is in multiples of whole numbers or fractional division; estimates readings between tick marks.

5 = E. Understands deductive reasoning.

1. Correctly interprets sentences that incorporate logical connectives, "and," "or," and "if-then" as well as the quantifiers "some," "all," and "none."

2. Uses deductive reasoning to determine whether a conclusion based on a series of statements about everyday situations is valid or invalid.

3. Sees the need for basic definitions and assumptions and recognizes hidden assumptions--e.g., in advertisements or political slogans.

4. Makes appropriate generalizations; identifies counterexamples to inappropriate generalizations; e.g., the product is not always larger than the number multiplied.

6 = F. Can interpret graphic, symbolic, and verbal material.

1. Makes reasonable visual comparisons of the sizes of two or more objects.

2. Reads and interprets bar, line, and circle graphs and pictographs.

3. Chooses a mathematically appropriate graph to represent a given set of data.

4. Interprets a schematic diagram--e.g., a flowchart, electrical wiring diagram, or diagram of the circulatory system of a frog.

SPECIFICATIONS (continued)

PICTURE STIMULUS

1 = Picture	A picture or diagram appears as a stimulus.
2 = Nonpicture	There is no picture or diagram as stimulus.

VARIABLES STUDIED IN GENERAL KNOWLEDGE: SCIENCE

(See also minority stimulus, gender reference in stimulus, negative item, and Roman numeral format in the section "Variables Common to All Tests.")

SKILLS/ABILITIES

1 = Knowledge	Questions in this category involve recall or recognition of material that a candidate has previously encountered. More complicated thought processes are unlikely to be helpful in selecting the correct answer. A candidate either recognizes the correct answer or misses the question.
2 = Comprehension	Questions classified here involve translation of, or extrapolation from, material presented in the stem of the question or in stimulus material provided with the question. For such a question, the candidate must be provided with something to translate. A statement, a symbolic equation, a graph, and a diagram are examples of the sorts of information that may be provided. The correct response, then, is a different way of expressing the same information. The stem of a question might include the statement of a law and the options might be possible examples of applications of the law, but the law must be stated if the question is to fit into this category.
3 = Application	Application questions require that the candidate apply the laws, principles, concepts, etc., that are being tested. The laws, etc., are not stated.
4 = Analysis or Synthesis	Like comprehension questions, questions in this category must be based on stimulus material provided with the question. This time the candidate must analyze the stimulus material rather than translate it in order to select the answer. The

APPLICATION (continued)

analysis may involve recognizing aspects of the stimulus (the main theme, the strongest argument, a conclusion, etc.), or the analysis may involve handling the content of the stimulus in an analytical/ critical way. Synthesis questions involve the putting together of two or more ideas not previously combined by the candidate. Requiring a candidate to go through, during his or her thinking, an intervening step between the statements in the stem and the statements that are options is one way that may lead to a synthesis question.

5 = Evaluation

Questions for this level involve two criteria for selecting the correct response, and both criteria must be used in the reasoning process. For example, a question might ask candidates for a safe way of preparing oxygen in a junior high classroom. At least two but not all of the options must be ways of combining oxygen. Then, from the choices that would produce oxygen, the candidate must choose the one that is safe for a junior high classroom.

SPECIFICATIONS

1 = 1

Energy relationships are important in both living and nonliving contexts.

- The Sun is the dominant source of all or nearly all of the Earth's energy.
- Energy transfer occurs in food webs and food chains.
- Energy changes are vital in all life processes of all organisms.
- Energy is conserved but can be converted from one form to another.

SPECIFICATIONS (continued)

2 = 2

Living things have significant features.

- Living things, in contrast to nonliving things, possess cellular structure.
- Living things have the following functions: food getting, digestion, transport, assimilation, respiration, excretion, response to environmental stimuli, and reproduction.

3 = 3

The operation of natural processes has resulted in organisms that fill a vast number of ecological niches, and these organisms are usually classified on a structural basis into a small number of categories that facilitate the understanding and study of the organisms.

- Processes at work have led to changes in the dominant organisms at various times and in various places.
- There is great variety among organisms, both plant and animal.
- Animals, and particularly humans, have developed systems of communicating among themselves.

4 = 4

There are significant relationships between living organisms, particularly humans, and the environment.

- Living organisms, and especially humans, can change their environments. The changes may be beneficial but are often destructive.
- Some resources are nonrenewable and others can be renewed.

SPECIFICATIONS (continued)

5 = 5

Earth is both a part of the universe and a body that has special characteristics.

- Various evidences provide insight into the origins of the universe and of the Earth.
- Earth has various components. The arrangements of some of these components have changed and are changing with time in response to various forces that produce these rearrangements.

6 = 6

All matter is composed of atoms, atoms are divisible, and atoms undergo combinations.

- Matter exhibits three states, and these states can be described in molecular terms.
- Atoms contain nuclei and are composed of smaller particles.
- Atoms combine in various ways, resulting in a wide diversity of substances.
- The properties of substances determine the applications of these substances: metals and nonmetals, elements and compounds, acids and bases, organic substances and inorganic substances, etc.

7 = 7

There are important forces that act on units of matter.

- Almost all interactions of matter result from the operation of a limited number of forces: gravitational, electromagnetic, and nuclear.

SPECIFICATIONS (continued)

- Humans and their inventions can manipulate forces to bring about important movements and changes.

8 = 8

There are characteristic methods that have contributed to the development of science. Included are kinds of reasoning and the organization of information.

- Humans can increase their understanding of the laws of science and can use them but cannot alter them.

- There is dynamic interaction between pure science and technology.

- Models are developed to coordinate and explain observations and results of experiments.

9 = 9

Science has a role in securing and maintaining important human values.

- Knowledge of nutrition and of the causes of illness and disease enables humans to employ strategies for health maintenance, illness prevention, and safety.

- Science has provided the means of maintaining life while making time available for a wide range of human activities.

PICTURE STIMULUS

1 = Picture

Graph, diagram, drawing, etc.

2 = Nonpicture

No graph, diagram, drawing, etc.

SUBJECT MATTER

1 = Biology	Material typically covered in biology courses prior to receiving a degree in education.
2 = Chemistry	Material typically covered in chemistry courses prior to receiving a degree in education.
3 = Physics	Material typically covered in physics courses prior to receiving a degree in education.
4 = Other	Material typically covered in astronomy, geology, and meteorology courses prior to receiving a degree in education.

VARIABLES STUDIED IN GENERAL KNOWLEDGE: SOCIAL STUDIES

(See also minority stimulus, gender reference in stimulus, negative item, and Roman numeral format in the section "Variables Common to All Tests.")

SETS VERSUS DISCRETES

1 = Discrete	A discrete is a single item based on a stimulus.
2 = Linked Set	A set is made up of one or more stimuli and two or more items.

SPECIFICATIONS (Themes)

1 = 1	Forces that have shaped human culture and institutions (e.g., industrialization, migration, war, revolution)
2 = 2	Understanding the behavior of individuals, small groups, and institutions (e.g., social roles, personality, social interaction)
3 = 3	Recognizing the universal features of world cultures and differences in cultures (e.g., language, social taboos, religion)
4 = 4	Essential tools and perspectives to make informed judgments about society (e.g., interpreting graphs, data, maps)

PICTURE STIMULUS

1 = Picture	Stimulus containing graphs, maps, tables, charts, photographs, political cartoons, etc.
2 = Nonpicture	Item does not contain graphs, maps, tables, etc.

CONTENT OBJECTIVES

1 = 1	U.S. historical and cultural events, political institutions and values (e.g., American Revolution, Civil War, Prohibition)
2 = 2	Prominent characteristics of societies and cultures (e.g., government, law, religion)
3 = 3	Relationship between culture and the individual (e.g., social roles and interaction)
4 = 4	Economic processes (e.g., basic understanding of the market and forces that shape the market)
5 = 5	Geographical features and characteristics of human settlement (e.g., land use patterns, changing climate, agricultural development; harbors, rivers, ports, transport, permanence)
6 = 6	Methodologies and methodological tools of the social sciences (e.g., analyzing data, graphs, and maps)

SKILLS

1 = Knowledge	Questions in this category involve recall or recognition of material that a candidate has previously encountered. More complicated thought processes are unlikely to be helpful in selecting the correct answer. A candidate either recognizes the correct answer or misses the question.
2 = All other skills	Questions that test skills other than knowledge or recall are categorized as 2.

VARIABLES STUDIED IN PROFESSIONAL KNOWLEDGE

(See also minority stimulus, gender reference in stimulus, negative item, and Roman number format in the section "Variables Common to All Tests.")

QUESTION FORMAT

1 = Open	The question stimulus is open-ended, and the options complete the sentence.
2 = Closed	The question stimulus is a closed stem; i.e., it ends with a colon or a question mark.

SPECIFICATIONS

Items may be classified in one or two categories. The specifications for the tests studied actually called for a number of questions to test both I and II.

1 = I. The Process of Teaching

The beginning teacher is expected to have knowledge about and utilize cognitive processes relevant to the process of teaching. The teacher should know

A. How to Plan

B. How to Implement Plans

C. How to Evaluate

D. How to Behave in a Professional Manner

2 = II. The Context of Teaching

The beginning teacher is expected to have knowledge about and utilize cognitive processes relevant to the context of teaching. The teacher should

A. Recognize the Constitutional Rights of Students and the Implications for Classroom Practice

B. Recognize the spirit and implications of State/Federal/Judicial Policy, including the spirit and implications of policy

SPECIFICATIONS (continued)

C. Recognize the Extraclassroom Influences on Teachers

D. Recognize the Extraclassroom Influences on Students

E. Demonstrate a Knowledge of One's Occupation/ Profession

3 = III. Both the Process and the Context of Teaching

RESPONSE LENGTH

1 = Complete sentence response

2 = two- to three-line response--not a sentence

3 = Short-phrase response (up to one line)

4 = Single/double-word response

GERUND RESPONSE

1 = If any option has a single gerund phrase response

2 = If any option has a two-part gerund phrase response

3 = N/A, no gerunds present

EVALUATION OF ACTION

1 = Question calls for a response about whether an action includes "inappropriate" or "appropriate."

2 = All other questions

QUESTION REFERENT

1 = Refers to teacher only

2 = Refers to teacher and student

3 = Refers to teacher and parent

4 = Refers to teacher/student and parent

5 = Refers to student only

6 = Refers to student and parent

QUESTION REFERENT (continued)

7 = Refers to Parent Only

8 = Nothing

OPTIONS ARE VERBAL RESPONSES

1 = The options are stated as quoted verbal responses.

2 = N/A

DOUBLE FACTOR

1 = Two or more pieces of information are required to arrive at the intended answer.

2 = Only one piece of information is needed to arrive at the intended answer.

ORDERING FORMAT

1 = Question asks examinees to put things in the correct order.

2 = No ordering is required.

PICTURE STIMULUS (Graphic)

1 = Picture	Stimulus contains a picture
2 = Graph	Stimulus contains a graph
3 = Table	Stimulus contains a table
4 = Nothing	No visual stimulus given

QUESTION BASED ON LEGAL KNOWLEDGE

1 = Legal Knowledge	Question asks about the Constitution, laws, federal regulations, state policies, etc., that affect education
2 = N/A	No legal knowledge required

PROFESSIONAL JUDGMENT/CONSENSUS QUESTION

1 = Based on consensus/judgment

2 = Based on fact/theory/research

COMPLEXITY (in terms of length of material)

1 = Complex stem/complex response

2 = Complex stem/easy response

3 = Easy stem/complex response

4 = Easy stem/easy response

CONTROVERSIAL TOPICS

1 = Controversial Topics — Topic is controversial, e.g., the Holocaust, evolution, sex education, religion, politics

2 = All Other Questions

THEORY/THEORIST QUESTIONS

1 = Question asks about a theory or theorist

2 = Question does not ask about a theory or theorist

Appendix B

Table B1

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Reading

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	6	-.26	.15	.25	.73	.10	.26
Hispanic	0	--	--	--	--	--	--
Other	7	-.25	.31	.15	.21	-.01	.31
General	32	.06	.35	-.02	.38	-.08	.42
Nothing	3	.15	.13	-.61	.40	.20	.10
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.04		.05		.54	
Gender Reference-Stimulus							
Female	2	.11	.02	.18	.14	-.44	.29
Male	4	-.03	.43	.04	.85	-.22	.50
Mixed	39	-.04	.36	.03	.38	-.01	.36
Neutral	3	.15	.13	-.61	.40	.20	.10
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.80		.13		.23	
Negative Item							
Negative Stem	1	-.33	--	1.05	--	.13	--
Positive Stem	47	-.01	.35	-.02	.44	-.03	.39
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.38		.02		.68	

Table B1 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Reading

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	3	.15	.19	.05	.32	.08	.33
Non-Roman	45	-.03	.36	.00	.47	-.04	.39
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.39		.86		.63	
Content							
Teacher-Related	28	.06	.37	-.04	.40	-.04	.41
General Interest	20	-.13	.29	.06	.53	-.01	.34
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.06		.46		.78	
Subject Matter							
Science/Technology	14	-.01	.26	-.20	.40	-.18	.38
Other	34	-.02	.38	.08	.46	.03	.37
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.94		.06		.08	
Question Format							
Closed	24	.03	.34	.04	.51	-.09	.35
Open	24	-.07	.36	-.05	.40	.03	.40
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.33		.52		.30	

Table B1 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Reading

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Passage Length							
Long Passage	18	-.03	.42	.05	.35	.06	.38
Short Passage	17	-.02	.33	-.03	.48	-.06	.38
Statements	13	.00	.25	-.04	.56	-.12	.36
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.96		.87		.43	
Sets Versus Discretes							
Set	38	-.05	.35	.02	.40	.04	.37
Discrete	10	.09	.32	-.10	.62	-.28	.32
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.29		.47		.02	
Named I-Minority							
No One Named	35	-.08	.30	-.04	.45	-.10	.35
Black Named	2	-.20	.24	.86	.30	.16	.42
Hispanic Named	0	--	--	--	--	--	--
Native American	0	--	--	--	--	--	--
Asian American	0	--	--	--	--	-	--
Other Minority	1	-.38	--	-.16	--	-.13	--
Nonminority	10	.28	.38	-.03	.38	.20	.40
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.02		.06		.14	
Named II-Gender							
No One Named	35	-.08	.30	-.04	.45	-.10	.35
Female Named	0	--	--	--	--	--	--
Male Named	5	-.01	.39	.09	.68	-.13	.37
Mixed Named	8	.26	.40	.09	.30	.36	.28
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.04		.70		.01	

Table B1 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Reading

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications I-Subskills							
Main Idea	2	-.02	.02	.03	.09	.04	.09
Supporting Idea/Detail	13	.07	.25	.17	.43	-.02	.39
Relationships	2	-.01	.01	.12	.12	-.15	.12
Paraphrase/ Summary	4	.11	.37	-.65	.43	.03	.22
Author's Purpose	6	-.08	.48	-.12	.18	.13	.41
Author's Assumption	1	-.10	--	.46	--	.20	--
Author's Attitude or Tone	2	.36	.12	-.19	.54	.62	.28
Implications/ Inferences	6	-.34	.33	.25	.46	-.04	.31
Fact Versus Opinion	0	--	--	--	--	--	--
Organization	4	-.09	.18	-.04	.24	-.18	.20
Use of Language	1	-.43	--	.56	--	-.26	--
Application	1	.04	--	-.34	--	-.29	--
Emotional/ Manipulative Aspects of the Passage	0	--	--	--	--	--	--
Strengths/ Weaknesses of the Argument	2	-.32	.07	-.58	.23	-.76	.07
Relevance/ Appropriateness of Supporting Arguments or Evidence	2	.58	.02	-.04	.37	-.33	.17
Relation of the Passage to the Audience	2	-.07	.28	.17	.42	.17	.34
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.16		.11		.15	

Table B1 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Reading

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications II							
Comprehension	21	.06	.26	.00	.50	-.01	.33
Analysis	21	-.13	.38	.04	.40	.03	.38
Evaluation	6	.07	.41	-.15	.47	-.31	.44
TOTAL	48	-.02	.35	.00	.46	-.03	.38
SIGNIFICANCE		.19		.67		.16	

Table B2

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Listening

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	0	--	--	--	--	--	--
Hispanic	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
General	57	.02	.59	.06	.55	-.01	.49
Nothing	4	-.02	.44	-.01	.39	-.58	.42
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		.91		.79		.03	
Gender Reference-Stimulus							
Female	13	.11	.60	.26	.48	.04	.56
Male	7	.40	.39	.17	.53	-.02	.36
Mixed	37	-.08	.58	-.03	.56	-.02	.48
Neutral	4	-.02	.44	-.01	.39	-.58	.42
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		.23		.42		.18	
Negative Item							
Negative Stem	0	--	--	--	--	--	--
Positive Stem	61	.02	.58	.06	.55	-.04	.50
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		N/A		N/A		N/A	

Table B2 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Listening

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	0	--	--	--	--	--	--
Non-Roman	61	.02	.58	.06	.55	-.04	.50
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		N/A		N/A		N/A	
Item Type							
Statements	26	-.08	.51	.11	.43	-.25	.54
Questions	8	.44	.32	.43	.38	.32	.49
Dialogues	15	.36	.40	-.13	.71	.09	.40
Mini-talks	12	-.49	.58	-.07	.48	.00	.32
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		.0001		.09		.02	
Speaker Gender							
Female Speaker	16	.20	.46	.27	.47	-.23	.61
Male Speaker	14	-.12	.46	.15	.42	.03	.54
Mixed	31	-.01	.66	-.09	.59	.02	.39
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		.30		.08		.23	

Table B2 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Listening

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications							
Basic Comprehension	27	-.04	.61	.01	.36	-.12	.46
Analysis	19	-.11	.60	.07	.61	-.18	.51
Evaluation	7	.07	.38	-.03	.67	.22	.32
Feedback	8	.44	.32	.27	.70	.33	.48
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		.16		.66		.04	
Specifications-Subskills							
IA-Noting Details	10	.19	.64	.03	.38	-.02	.39
IB-Paraphrasing	10	-.11	.48	.14	.28	-.26	.60
IC-Recognizing Cognitive Clues	2	.47	.06	.07	.24	-.35	.09
ID-Understanding Connotations of Words	0	--	--	--	--	--	--
IE-Interpreting Figurative Language	1	-.02	--	-.31	--	-.07	--
IF-Identifying the Main Idea	4	-.70	.43	-.32	.31	.05	.19
IIA-Identifying Assumptions Made by the Speaker	0	--	--	--	--	--	--
IIB-Drawing Inferences From the Information Contained in the Message	10	.33	.56	.00	.56	-.26	.46
IIC-Recognizing the Implications of Assertions Made in the Message	7	.04	.57	-.04	.61	-.27	.49
IID-Identifying the Speaker's Tone	2	.52	.21	.82	.26	.53	.27

Table B2 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Listening

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications-Subskills, continued							
IIIA-Identifying and Evaluating How a Persuasive Argument is Constructed	6	-.03	.32	-.11	.70	.22	.34
IIIB-Assessing the Appropriateness of Supporting Material	0	--	--	--	--	--	--
IIIC-Evaluating the Effect of the Speaker's Tone on the Audience	1	.66	--	.40	--	.20	--
IVA-Feedback-Response	8	.44	.32	.27	.70	.33	.48
TOTAL	61	.02	.58	.06	.55	-.04	.50
SIGNIFICANCE		.03		.56		.14	

Table B3

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Writing

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	4	.02	.35	.19	.66	-.09	.31
Hispanic	1	-1.35	--	-.19	--	.51	--
Other	3	.25	.36	.00	.68	.27	.32
General	44	.11	.48	-.02	.60	-.06	.55
Nothing	17	-.15	.35	.23	.67	.02	.51
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		.01		.70		.70	
Gender Reference-Stimulus							
Female	4	.16	.08	.10	.53	-.32	.43
Male	13	-.08	.40	-.05	.50	-.13	.63
Mixed	34	.15	.54	-.04	.62	.04	.49
Neutral	18	-.17	.35	.29	.69	.01	.50
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		.10		.33		.52	
Negative Item							
Negative Stem	0	--	--	--	--	--	--
Positive Stem	69	.02	.48	.05	.63	-.02	.53
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		N/A		N/A		N/A	

Table B3 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Writing

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	0	--	--	--	--	--	--
Non-Roman	69	.02	.48	.05	.63	-.02	.53
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		N/A		N/A		N/A	
Item Type							
Usage	39	.09	.53	.30	.61	.14	.51
Sentence Correction	30	-.07	.38	-.26	.51	-.22	.48
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		.17		.0002		.0049	
Content of Stimulus							
Humanities	48	.05	.50	.09	.62	.01	.48
Science	21	-.02	.43	-.03	.64	-.09	.61
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		.59		.50		.44	

Table B3 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Writing

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Underline Position-Sentence Correction							
Beginning of Sentence	4	-.37	.12	-.16	.48	-.28	.32
Middle of Sentence	19	-.03	.38	-.21	.53	-.23	.47
End of Sentence	5	-.04	.49	-.53	.42	-.34	.59
Entire Sentence	2	.16	.13	-.25	.17	.29	.13
TOTAL	30	-.07	.38	-.26	.51	-.22	.48
SIGNIFICANCE		.39		.67		.48	
Key-Combined							
Error	55	.07	.51	.05	.67	.00	.57
No Error	14	-.16	.26	.08	.47	-.07	.33
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		.11		.85		.67	
Key-Usage							
Error	32	.17	.55	.31	.67	.15	.55
No Error	7	-.23	.21	.21	.19	.08	.22
TOTAL	39	.09	.53	.30	.61	.14	.51
SIGNIFICANCE		.07		.68		.76	
Key-Sentence Correction							
Error	23	-.06	.41	-.33	.45	-.22	.51
No Error	7	-.09	.29	-.04	.62	-.23	.35
TOTAL	30	-.07	.38	-.26	.51	-.22	.48
SIGNIFICANCE		.88		.20		.97	

Table B3 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Communication Skills: Writing

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications							
Preposition	0	--	--	--	--	--	--
Noun	2	.16	.07	-.07	.15	.25	.02
Pronoun	6	.55	.34	.54	.38	.21	.66
Verb	10	.13	.60	.43	.81	.07	.48
Adjective	4	-.25	.35	.19	.41	-.14	.39
Adverb	2	-.27	.30	-.61	.61	-.85	.68
Coordination	2	.21	.34	-.36	.06	.36	.32
Subordination	2	-.20	.30	-.54	.30	-.33	.00
Correlation	1	.08	--	.82	--	.87	--
Comparison	2	.13	.03	-.00	.53	-.29	.20
Parallel	3	.39	.13	-.54	.34	-.01	.29
Negation	5	.34	.41	-.13	.53	-.17	.60
Modification	3	-.05	.31	-.23	.14	.10	.29
Diction and Word Choice	3	-.36	.34	-.12	.89	-.41	.63
Idiom	4	-.26	.65	.17	.40	.24	.62
Redundancy	5	-.25	.37	-.30	.60	-.06	.52
Punctuation	1	.81	--	.76	--	.29	--
No Error	14	-.16	.26	.08	.47	-.07	.33
TOTAL	69	.02	.48	.05	.63	-.02	.53
SIGNIFICANCE		.06		.19		.50	

Table B4

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Social Studies

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	3	.06	.39	1.03	.79	-.27	.63
Hispanic	0	--	--	--	--	--	--
Other	4	-.32	.60	.19	.17	.40	.68
General	18	.17	.60	-.10	.84	-.02	.63
Nothing	22	-.09	.54	-.19	.65	-.02	.53
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.36		.07		.52	
Gender Reference-Stimulus							
Female	1	1.03	--	-.37	--	-.66	--
Male	3	-.19	.23	.66	1.09	-.22	.26
Mixed	21	.07	.62	.02	.79	.09	.69
Neutral	22	-.09	.54	-.19	.65	-.02	.53
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.25		.34		.59	
Negative Item							
Negative Stem	6	.14	.80	-.31	.63	-.23	.65
Positive Stem	41	-.02	.54	-.01	.79	.04	.59
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.54		.38		.33	

Table B4 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Social Studies

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	4	.23	.26	-.08	.17	.02	.56
Non-Roman	43	-.02	.60	-.04	.81	.00	.61
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.42		.92		.96	
Item Type							
Discrete	36	.09	.56	.09	.65	-.02	.56
Linked Set	11	-.28	.57	-.50	.97	.07	.72
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.07		.03		.67	
Picture Stimulus							
Picture	14	-.08	.60	-.45	.96	-.28	.54
Nonpicture	33	.03	.57	.12	.61	.12	.59
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.54		.02		.04	

Table B4 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Social Studies

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Content Objectives							
U.S. Historical and Cultural Events, Political Institutions, and Values	11	-.04	.55	.56	.76	.14	.63
Prominent Characteristics of Societies and Cultures	8	.24	.39	-.14	.35	-.29	.51
Relationship Between Culture and the Individual	3	.71	.34	-.13	.49	.48	.46
Economic Processes	5	.02	.13	.03	.35	.35	.59
Geographical Features and Characteristics of Human Settlement	6	-.44	.60	-.09	.44	.12	.48
Methodologies and Methodological Tools of the Social Sciences	14	-.08	.63	-.46	.92	-.22	.54
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.08		.05		.15	
Skills							
Knowledge	11	-.08	.66	.03	.73	-.03	.53
Other	36	.02	.55	-.07	.79	.01	.62
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.62		.70		.87	

Table B4 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Social Studies

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications							
Forces That Have Shaped Human Culture and Institutions	10	-.32	.31	-.34	.78	-.13	.43
Understanding the Behavior of Individuals, Small Groups, and Institutions	14	.22	.47	.25	.73	.18	.69
Recognizing the Universal Features of World Cultures and Differences in Cultures	13	-.17	.59	.13	.60	.14	.56
Essential Tools and Perspectives to Make Informed Judgments about Society	10	.22	.69	-.39	.79	-.31	.54
TOTAL	47	.00	.58	-.05	.78	.00	.60
SIGNIFICANCE		.06		.11		.18	

Table B5

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Mathematics

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	0	--	--	--	--	--	--
Hispanic	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
General	13	-.12	.68	.02	.50	-.08	.50
Nothing	27	.01	.64	-.01	.47	.03	.47
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		.57		.88		.51	
Gender Reference-Stimulus							
Female	1	-1.80	--	.74	--	.11	--
Male	1	.31	--	.17	--	.72	--
Mixed	11	.00	.50	-.06	.49	-.17	.47
Neutral	27	.01	.64	-.01	.47	.03	.47
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		.05		.47		.31	
Negative Item							
Negative Stem	7	.17	.54	.09	.68	-.12	.48
Positive Stem	33	-.07	.67	-.02	.42	.03	.48
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		.40		.60		.47	

Table B5 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Mathematics

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	0	--	--	--	--	--	--
Non-Roman	40	-.03	.66	.00	.48	.00	.48
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		N/A		N/A		N/A	
Picture Stimulus							
Picture	15	-.04	.45	-.25	.39	.00	.41
Nonpicture	25	-.02	.76	.15	.46	-.01	.52
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		.93		.01		.95	
Specifications							
A-Has Good Number Sense, That is, Understands How Numbers Behave	8	.24	.55	.06	.35	-.02	.50
B-Understands and Uses Numbers in an Appropriate Way to Quantify Thinking	8	-.37	.82	.26	.65	-.07	.57
C-Recognizes and Uses Mathematical Relationships	9	.07	.76	-.11	.30	.01	.44
D-Understands the Mathematical Basis of Measurement	5	-.20	.46	-.29	.41	.23	.22
E-Understands Deductive Reasoning	5	.19	.40	.28	.25	-.09	.59
F-Can Interpret Graphic, Symbolic, and Verbal Material	5	-.13	.32	-.31	.42	-.02	.37
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		.50		.13		.93	

Table B5 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Mathematics

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Item Content							
Spatial Only	6	.07	.56	-.35	.40	-.03	.50
Inferential	13	.07	.85	.22	.39	-.05	.53
Estimation	1	.76	--	.33	--	.27	--
Spatial and Inferential	3	-.25	.26	-.40	.50	-.05	.44
Spatial and Estimation	4	-.40	.23	-.31	.30	.01	.52
Other Items	13	-.07	.57	.11	.45	.04	.43
TOTAL	40	-.03	.66	.00	.48	.00	.48
SIGNIFICANCE		.66		.05		.99	

Table B6

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Science

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	0	--	--	--	--	--	--
Hispanic	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
General	12	.22	.60	.13	.64	-.03	.54
Nothing	36	-.08	.71	.03	.50	.01	.48
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.20		.55		.78	
Gender Reference-Stimulus							
Female	3	.04	.04	.69	.06	.34	.25
Male	0	--	--	--	--	--	--
Mixed	9	.29	.68	-.05	.65	-.16	.56
Neutral	36	-.08	.71	.03	.50	.01	.48
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.39		.11		.33	
Negative Item							
Negative Stem	6	.30	.54	.14	.59	-.09	.34
Positive Stem	42	-.04	.70	.04	.53	.01	.52
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.26		.67		.66	

Table B6 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Science

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	4	.16	.44	-.26	.42	-.24	.27
Non-Roman	44	-.02	.71	.08	.54	.02	.51
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.64		.23		.33	
Skills/Abilities							
Knowledge	14	-.14	.63	.04	.51	-.22	.58
Comprehension	2	.03	.11	-.26	.46	-.14	.10
Application	30	.11	.72	.06	.57	.11	.45
Analysis or Synthesis	0	--	--	--	--	--	--
Evaluation	2	-.66	.58	.31	.07	.09	.21
TOTAL	48	.00	.69	.52	.54	.00	.50
SIGNIFICANCE		.39		.79		.25	
Picture Stimulus							
Picture	10	.18	.61	-.18	.46	-.10	.32
Nonpicture	38	-.05	.71	.11	.54	.03	.53
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.37		.14		.49	
Subject Matter							
Biology	23	.28	.68	.20	.51	.06	.54
Chemistry	6	-.35	.35	-.04	.27	-.18	.37
Physics	5	-.44	.91	-.07	.65	-.01	.28
Other	14	-.15	.52	-.11	.56	-.01	.51
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.05		.35		.80	

Table B6 (continued)

Means and Standard Deviations of Categories
for Variables on NTE General Knowledge: Science

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications							
Energy Relationships	6	-.39	.61	-.07	.42	.01	.30
Living Things...	5	.47	.64	.08	.54	.11	.56
The Operation of Natural Processes	5	.14	.24	.40	.31	.13	.34
Significant Relationships Between Living Organisms...	5	-.27	.30	-.19	.59	.02	.80
Earth...	4	-.09	.51	-.10	.20	-.22	.78
Atoms...	6	-.19	.19	-.05	.26	-.30	.16
Matter...	5	-.60	.89	.02	.70	.11	.36
Scientific Methods...	6	-.06	.55	.26	.53	.31	.14
Science's Role...	6	.91	.49	.08	.72	-.17	.41
TOTAL	48	.00	.69	.05	.54	.00	.50
SIGNIFICANCE		.0044		.81		.61	

Table B7

Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Literature and Fine Arts

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	2	-.20	.11	1.60	2.20	.11	.11
Hispanic	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
General	33	.09	.56	-.01	.58	.04	.53
Nothing	19	-.15	.55	-.07	.49	-.01	.36
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.29		.01		.91	
Gender Reference-Stimulus							
Female	10	-.03	.52	.24	.57	.30	.65
Male	13	.08	.59	-.15	.62	.00	.40
Mixed	12	.15	.52	.20	1.15	-.13	.40
Neutral	19	-.15	.55	-.07	.49	-.01	.36
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.48		.49		.18	
Negative Item							
Negative Stem	7	.01	.49	-.24	.74	-.03	.35
Positive Stem	47	-.01	.57	.07	.74	.03	.48
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.92		.31		.77	

Table B7 (continued)

Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Literature and Fine Arts

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	5	.09	.40	-.41	.27	-.09	.37
Non-Roman	49	-.02	.57	.07	.77	.03	.48
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.68		.18		.57	
Sets Versus Discretes							
Linked Sets	23	.03	.43	-.04	.52	.07	.33
Discrete	31	-.03	.64	.08	.88	-.01	.55
TOTAL	54	.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.69		.59		.55	
Single versus Multiple Stimuli							
Single	46	.00	.57	.04	.77	.01	.48
Multiple	8	-.04	.50	-.02	.61	.08	.37
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.88		.85		.69	
Picture Stimulus							
Picture	26	-.06	.61	-.04	.65	.12	.54
Nonpicture	28	.04	.50	.09	.82	-.07	.37
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.51		.54		.15	

Table B7 (continued)

Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Literature and Fine Arts

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Skills							
Identification	8	-.05	.51	.11	.62	.11	.37
Analysis and Interpretation	42	.07	.51	-.10	.50	-.06	.38
Comparison	4	-.74	.59	1.18	1.63	.71	.79
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.02		.0035		.0047	
Discipline							
Fiction	12	.06	.49	.22	1.16	-.14	.42
Poetry	13	-.04	.33	.01	.34	.09	.24
Painting	9	-.02	.47	-.07	.53	.05	.38
Sculpture	6	.07	.62	.39	.80	.42	.75
Architecture	8	-.59	.45	-.01	.34	.13	.41
Performing Arts	6	.66	.45	-.48	.64	-.39	.30
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.0016		.43		.04	
Culture							
American	13	-.08	.40	.38	1.12	-.02	.26
European	32	.10	.47	-.05	.52	-.04	.39
Other	9	-.28	.86	-.19	.59	.31	.75
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.18		.14		.13	
Time Period							
Pre-1800	17	-.02	.63	-.13	.42	-.02	.40
1800	17	-.09	.41	.11	.51	.11	.33
1900	20	.08	.59	.09	1.05	-.02	.60
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.66		.60		.67	

Table B7 (continued)

Means and Standard Deviations of Categories for Variables on NTE General Knowledge: Literature and Fine Arts

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Specifications							
I-The Ability to Recognize Basic Elements and Components of Works of Literature and Fine Art	10	-.10	.59	-.19	.41	-.04	.44
II-The Ability to Analyze and Interpret Works of Literature and Fine Art by Dealing with the Interrelation-ships of the Elements of Works	34	.08	.46	-.06	.53	-.07	.37
III-The Ability to Relate Works of Literature and Fine Art to One Another and to Their Social/ Historical Contexts	10	-.21	.75	.55	1.23	.40	.59
TOTAL	54	-.01	.56	.03	.75	.02	.47
SIGNIFICANCE		.31		.04		.02	

Table B8

Means and Standard Deviations of Categories
for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Minority Reference-Stimulus							
Black	1	-.08	--	1.36	--	-.03	--
Hispanic	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
General	141	.12	.53	-.02	.46	.03	.55
Nothing	33	-.39	.71	.01	.59	-.11	.47
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		<.0000		.02		.43	
Gender Reference-Stimulus							
Female	4	-.01	.41	-.15	.29	.18	.36
Male	3	.09	.62	-.15	.23	.45	.29
Mixed	135	.12	.54	.00	.48	.01	.56
Neutral	33	-.39	.71	.01	.59	-.11	.47
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.0003		.89		.27	
Negative Item							
Negative Stem	23	.21	.45	-.20	.43	-.09	.49
Positive Stem	152	-.01	.62	.02	.50	.02	.55
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.10		.04		.40	

Table B8 (continued)

Means and Standard Deviations of Categories for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Roman Numeral Format							
Roman	6	.23	.42	.17	.40	.34	.29
Non-Roman	169	.01	.61	-.01	.50	-.01	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.39		.39		.12	
Question Format							
Open-Ended	68	-.10	.58	.02	.53	-.04	.58
Closed Stem	107	.09	.61	-.02	.47	.03	.51
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.03		.55		.41	
Specifications							
Category I Only-The Process of Teaching	50	.21	.57	.01	.42	-.02	.48
Category II Only-The Context of Teaching	38	-.33	.60	.10	.55	-.05	.63
Combined I & II	87	.06	.56	-.06	.50	.04	.53
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.0002		.21		.67	

Table B8 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Response Length							
Complete Sentence Response	41	.12	.39	.04	.41	-.02	.54
Two to Three Lines --Not a Sentence	80	.05	.55	.05	.54	.07	.52
Short Phrase Response	39	-.11	.77	-.12	.44	-.10	.55
Single/Double Word Response	15	-.11	.77	-.15	.54	-.05	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.30		.17		.42	
Gerund Response							
If Any Option Has a Single Gerund Phrase Response	33	.09	.52	-.14	.35	.06	.40
If Any Option Has a Two-Part Gerund Phrase Response	11	.30	.55	.31	.63	.18	.56
Not Applicable, No Gerunds Present	131	-.02	.62	.00	.50	-.03	.56
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.18		.03		.39	

Table B8 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Evaluation of Action							
Question Calls for a Response About Whether an Action Includes "Inappropriate" or "Appropriate"	2	.12	.07	.65	.19	.76	.60
All Other Questions	173	.02	.61	-.01	.49	-.01	.53
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.81		.06		.05	
Question Referent							
Teacher Only	31	.08	.52	-.02	.57	.10	.47
Teacher and Student	37	.13	.51	-.01	.41	.06	.60
Teacher and Parent	4	.44	.23	-.11	.40	-.45	.47
Teacher, Student, and Parent	11	.50	.49	-.10	.55	-.33	.45
Student Only	42	.10	.52	-.01	.40	.06	.56
Student and Parent	2	-.10	.06	-.29	.21	-.36	.05
Parent Only	0	--	--	--	--	--	--
Nothing	48	-.32	.67	.05	.57	-.03	.50
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.0001		.93		.13	

Table B8 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Options Are Verbal Responses							
The Options Are Stated as Quoted Verbal Responses	5	.52	.33	.15	.39	-.20	.50
Not Applicable	170	.00	.60	-.01	.50	.01	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.06		.48		.40	
Double Factor							
Two or More Pieces of Information Are Required to Arrive at the Intended Answer	34	.20	.43	-.07	.45	.04	.49
Only One Piece of Information Is Needed to Arrive at the Intended Answer	141	-.03	.63	.01	.50	-.01	.55
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.04		.37		.63	
Ordering Format							
Question Asks Examinees to Put Things in the Correct Order	0	--	--	--	--	--	--
No Ordering is Required	175	.02	.61	-.01	.50	.00	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		N/A		N/A		N/A	

Table B8 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Picture Stimulus							
Picture	0	--	--	--	--	--	--
Graph	0	--	--	--	--	--	--
Table	1	.61	--	.67	--	-.44	--
Nothing	174	.01	.61	-.01	.49	.00	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.33		.17		.41	
Based on Legal Knowledge							
Require Knowledge of Legal Factors	14	-.64	.60	.15	.65	-.16	.51
Not Applicable	161	.07	.57	-.02	.48	.02	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		< .0000		.23		.24	
Professional Judgment/Consensus							
Based on Consensus/ Judgment	136	.06	.60	-.02	.49	.01	.55
Based on Fact/ Theory/Research	39	-.15	.60	.05	.51	-.01	.50
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.05		.41		.88	

Table B8 (continued)

Means and Standard Deviations of Categories
for Variables on NTE Professional Knowledge

Variable/Categories	Number of Items	Comparison					
		F/M		B/W		H/W	
		Mean	SD	Mean	SD	Mean	SD
Complexity-Length of Material							
Complex Stem/ Complex Response	37	.12	.48	.02	.41	-.03	.50
Complex Stem/Easy Response	12	.12	.35	-.22	.27	-.12	.48
Easy Stem/ Complex Response	19	.01	.52	.16	.60	.19	.45
Easy Stem/Easy Response	107	-.03	.67	-.02	.51	-.01	.57
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.61		.20		.39	
Controversial Topics							
Controversial Topics	22	-.38	.63	.04	.75	.00	.47
All Other Questions	153	.07	.58	-.01	.45	.00	.55
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.0010		.63		.99	
Theory/Theorist Questions							
Question Asks About a Theory or Theorist	5	-.09	.22	-.21	.36	-.09	.44
Question Does Not Ask About a Theory or Theorist	170	.02	.61	.00	.50	.01	.54
TOTAL	175	.02	.61	-.01	.50	.00	.54
SIGNIFICANCE		.70		.35		.69	